

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

Number of Countries: 001 Number of Patents: 001

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
CN 1239103	A	19991222	CN 98102506	A	19980617	200031 B

Priority Applications (No Type Date): CN 98102506 A 19980617

Patent Details:

Patent No	Kind	Lan Pg	Main IPC	Filing Notes
CN 1239103	A	1	C07K-014/765	

Abstract (Basic): CN 1239103 A

A process for producing human serum albumin by recombinant DNA technique, especially to the synthetic human serum albumin gene composed of yeast biased sex codons, is new.

INDEPENDENT CLAIMS are also included for the following: a recombinant expression vector, yeast host cells carrying the recombinant expression vector of the gene and obtained by transforming said vector, and the process for producing human serum albumin with said yeast host, especially the methyl nutritive yeast host in secretory mode.

Dwg.0/0

Title Terms: PROCESS; PREPARATION; RECOMBINATION; HUMAN; SERUM; ALBUMIN; COMPRISE; YEAST; BIAS; SEX; RECOMBINATION; DNA; TECHNIQUE

Derwent Class: B04; D16

International Patent Class (Main): C07K-014/765

International Patent Class (Additional): C12N-001/19; C12N-015/14; C12N-015/65

File Segment: CPI

60

[19]中华人民共和国国家知识产权局

[51]Int. Cl⁶

[12] 发明专利申请公开说明书

[21] 申请号 98102506.4

C07K 14/765
C12N 15/14 C12N 15/65
C12N 1/19
//C12N1/19, C12R1:
84

[43]公开日 1999 年 12 月 22 日

[11]公开号 CN 1239103A

[22]申请日 98.6.17 [21]申请号 98102506.4
[71]申请人 上海海济生物工程有限公司
地址 200031 上海市肇嘉浜路 288 号鼎新大楼 20 楼
[72]发明人 李绍极 陆德如

权利要求书 2 页 说明书 27 页 附图页数 14 页

[54]发明名称 重组人血清白蛋白的生产方法

[57]摘要

本发明涉及以重组 DNA 技术生产人血清白蛋白的方法,特别是涉及合成的由 酵母偏性密码子组成的并经过适当修饰的人血清白蛋白基因,携带所说的基因的重组表达载体和用所说的载体转化的酵母宿主细胞,以及由所说的酵母宿主,特别是甲基营养型酵母宿主以可分泌形式生产人血清白蛋白的方法。

ISSN 1008-4274

专利文献出版社出版

权 利 要 求 书

1. 基本上由酵母偏性密码组成的, 并经过适当修饰的合成的编码人血清白蛋白的核苷酸序列及其功能等同物。
2. 根据权利要求 1 的核苷酸序列, 其中对所说的修饰包括对编码人血清白蛋白的核苷酸序列的内含子剪接位点、非转录终止区的转录终止序列、长的重复序列、内部启动子序列及核糖体结合位点的修饰。
3. 根据权利要求 1 的核苷酸序列, 其中所说的编码人血清白蛋白的核苷酸序列是分别合成三组各 8 个核苷酸链, 然后再以基于聚合酶链反应的合成技术拼接成三个核苷酸片段, 再经克隆和纠正个别碱基错误后彼此串联而得到的。
4. 根据权利要求 1 的核苷酸序列, 其中对所说的人血清白蛋白编码序列具有如图 SEQ ID NO: 1 所示的核苷酸序列。
5. 包含根据权利要求 1 至 4 中任何一项的编码人血清白蛋白的核苷酸序列的重组表达载体。
6. 根据权利要求 5 的重组表达载体, 其中包括一个或多个由 (1) 甲基营养酵母之甲醇反应性基因的启动子区域, (2) 编码啤酒酵母 α 交配因子 (AMF) 前原序列和编码 HSA 的核苷酸序列, (3) 在甲基营养型酵母中有功能活性的转录终止子的, 以及 (4) 至少一个可选择标志基因和细菌复制原点组成的表达盒。
7. 用根据权利要求 5 或 6 的重组表达载体转化的甲基营养型酵母细胞。
8. 根据权利要求 7 的甲基营养型酵母细胞, 其为巴斯德毕赤酵母细胞。
9. 根据权利要求 7 的甲基营养型酵母细胞, 其中所说的细胞是具有 Mut⁺ 或 Mut⁻ 表型的。
10. 以 DNA 重组技术生产人血清白蛋白的方法, 该方法包括在适于表达所需的蛋白质的条件下培养按上述方法制备的包含编码人血清白蛋白之核苷

18.05.17

酸序列的酵母宿主细胞，然后从培养物中回收并纯化所需的人血清白蛋白。

说明书

重组人血清白蛋白的生产方法

本发明涉及以重组 DNA 技术生产人血清白蛋白的方法，特别是涉及合成的由酵母偏性密码子组成的并经过适当修饰的人血清白蛋白基因，携带所说的基因的重组表达载体和用所说的载体转化的酵母宿主细胞，以及由所说的酵母宿主，特别是甲基营养型酵母宿主以可分泌形式生产人血清白蛋白的方法。

人血清白蛋白（HSA）是人血浆的主要组成性蛋白质，其含量约占血浆总蛋白质的 50 % 以上。在人的正常生理条件下，人血浆白蛋白具有维持血浆胶体渗透压、营养和促进伤口愈合的作用，而且其作为载体物质，参与诸多疏水性生物分子如激素、生物学活性物质及药物等在血液中的运输。基于人血清白蛋白的这些重要生理学功能，该蛋白质在临床上具有十分重要的治疗应用价值，可用于治疗各种以血浆白蛋白丢失或白蛋白合成减少为基本特征之一的疾病，例如治疗各种创伤、烧伤、外科手术过程中或手术后失血，以及许多慢性消耗性疾病。

迄今为止，临床上使用的 HSA 和相关制剂主要是以供血者的血液或胎盘组织为来源，经提取和分离制备得到的。使用天然提取的 HSA 的主要缺点在于，一方面受到血液有限供应的限制，即有限的血液来源难以满足大量生产 HSA 和其相关制品的需要；另一方面，由于对供血者选择不当，常常可能造成血源污染，例如，肝炎病毒，甚至爱滋病毒的污染。因此，人们一直在试图建立商品化生产 HSA 可替代方法。随着重组 DNA 技术的进展，目前已有可能用这一技术大批量生产 HSA。

Lawn 等人（Nucleic Acids Research 9:6105, (1981)）和 Dugaiczky 等人（Proc. Natl. Acad. Sci. USA 79:71, (1982)）分别分离了编码人血清白蛋白的基因并测定了其核苷酸序列。根据所测得的核苷酸序列推断的 HSA 氨基酸序列，成熟的 HSA 是由 585 个氨基酸组成的，分子量约 66000 道尔顿，等电点为 $pI 4.7$ 的单链非糖基化蛋白质。该蛋白质最初产生时的是以前原蛋白质的形式存在的，在由细胞分泌和加工过程中分别除去信号肽和前原序列，得到成熟的 HSA 分子。成熟的 HSA 分子呈椭圆形三维空间结构，且含有多个二硫键。在分离得到 HSA 的 DNA 编码序列后，许多研究人员

试图利用各种不同的表达系统中以重组 DNA 技术大批量生产有相对较高纯度的 HSA。例如, Barns 描述了从正常人肝细胞中分离 HSA 的 mRNA, 并使用该 mRNA 作为模板, 借助反转录酶、Klenow 片段及 SI 核酸酶合成编码前 HSA 的全长 cDNA (参见欧洲专利申请 0206733), 并描述了携带该 cDNA 的质粒及其在大肠杆菌中的表达。然而, 由于人血清白蛋白的分子量较大, 构型复杂, 所以使用原核生物大肠杆菌 (Latta, M. et al., Bio/Technology, 5:1309-1314, (1987)) 和枯草芽孢杆菌 (Saunders, C.W. et al., J. Bacteriol. 169:2917-2925, (1987)) 等原核表达系统很难得到正确折叠的表达产物。另外, 由于在分泌表达产物往往存在难以去除的菌体内毒素及其他蛋白质污染物, 所以也很难以得到适于以大剂量临床应用的 HSA 产物。

使用哺乳动物细胞如 CHO 细胞, COS 细胞及昆虫细胞表达系统, 虽然能产生有正确空间构型的蛋白质产物, 但培养这些细胞的营养条件要求较高, 其传代时间远比微生物长, 以致须经过长时间培养才能达到足够高的细胞浓度, 而且一般情况下也难以对这些细胞进行必要的改良。因此, 从经济角度考虑, 使用这些真核生物体生产 HSA 是不适宜的。

因为所有真核生物体都具有表达遗传信息的机制, 所以可望在真核宿主中能够比在原核生物体 (例如大肠杆菌) 中更为有效地表达真核基因。近年来, 一些实验室使用低等真核生物体, 特别是酵母表达人血清白蛋白已取得了突破性进展。例如欧洲专利申请 0344459 描述了构建含有转录单元的质粒, 将此携带处于 AOX1 启动子控制下之 HSA 基因的质粒导入适当宿主, 特别是毕赤氏酵母菌株 (GTS115) 的 AOX1 基因区以得到转化体细胞, 然后在含有适当浓度 (0.1-5%) 甲醇的培养基中培养该转化体细胞, 得到以甲醇为唯一碳源的突变体菌株并在培养基分泌产生 HSA。有人对毕赤氏酵母表达产生的重组 HSA 的理化性质、空间结构及药理学特性与天然 HSA 进行了一系列比较实验, 结果表明两者几乎完全相同 (参见 Kazuo Ikegaya, Anal. Chem. 69:1986-1991, (1997))。另外, 欧洲专利申请 0339455、0123544、0248637 和 0251744 也分别公开了利用酵母宿主表达人血清白蛋白的方法及相应的产物纯化方法。因此, 这些研究成果使得人们利用重组 DNA 技术以高产量生产高纯度 HSA 成为可能。

为了进一步提高 HSA 在酵母细胞中的表达产率, 一些实验室还对 HSA 的信号肽、启动子序列, 以及 HSA 基因的拷贝数等进行了一系列深入研究。例如, 欧洲专利申请 0329127 和 0319641 分别公开了合成的 HSA 信号肽核苷酸编码序列, 将这些信号肽核苷酸编码序列连接到启动子 (例如 AOX1、AOX2、或 GAL1, 10) 序列与编码前原 HSA 的 DNA 序列之间, 以提高酵

母宿主表达和分泌HSA的能力。欧洲专利申请0506040公开了在天然AOX2启动子碱基序列中造成部分缺失或取代，或在其中加入新的碱基序列所得到的突变体AOX2启动子。

另一方面，在努力提高酵母细胞或其他宿主系统表达效力的同时，还着重就重组产生的HSA的纯化方法进行了深入地研究。例如，在培养HSA生产菌方面，还公开了向培养基中添加脂肪酸（欧洲专利申请0504823）、葡萄糖（日本专利公开JP - A219561/1989）、组氨酸（美国专利5612197）以提高生产菌的生长速率并改善HSA产率的方法；在乙酰色氨酸（1-100mM）或其他羧酸盐存在下，加热（50-70℃）处理含HSA的培养物上清以纯化HSA的方法（美国专利5132404）；联合使用超滤、加热处理、酸处理和超滤，然后再用阳离子交换剂、疏水层析载体和阴离子交换剂处理，并经盐析得到纯的HSA的方法（美国专利5440018）。另外，还公开了在胍类化合物的存在下培养表达HSA的重组体宿主细胞，以抑制重组HSA（rHSA）产物的颜色生成的方法（美国专利5369020）；以及用还原剂（如巯基化合物），或用阳离子交换剂除去游离多糖，然后再经热处理以使rHSA脱色的方法（欧洲专利申请0658569）。

如前所述，酵母细胞表达系统，特别是甲基营养型酵母表达系统是用于大批量发酵生产HSA的优选表达系统。尽管现有技术中针对重组载体的构建、用于基因表达的调控元件的修饰及重组体细胞的培养条件等作了多方面的改良，但迄今尚未见到的有关对HSA编码序列进行必要的修饰，通过改变HSA基因结构以进一步提高其在甲基营养型酵母中表达效率和水平的报导。

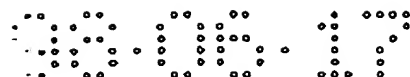
因此，本发明的一个目的是提供基本上由酵母偏性密码子组成并经过适当修饰的合成的编码人血清白蛋白的核苷酸序列及其功能等同物。

根据本发明这一目的的一个优选实施方案，其中对所说的修饰包括对HSA编码序列的内含子剪接位点、非转录终止区的转录终止序列、长的重复序列、内部启动子序列及核糖体结合位点的修饰。

根据本发明这一目的的一个优选方案，其中说的HSA编码序列是经分别合成三组各8个核苷酸链，然后以基于聚合酶链反应的合成技术拼接成三个核苷酸片段，再经克隆和纠正个别碱基错误后彼此串联连接而得到的。

根据本发明这一目的的一个优选方案，其中所说的HSA编码序列具有如图1A（SEQ ID NO 1）所示的核苷酸序列。

本发明的另一个目的是提供包含上述HSA编码序列的重组表达载体。



根据本发明这一目的的一个优选方案，其中所说的重组表达载体包括一个或多个由（1）甲基营养酵母之甲醇反应性基因的启动子区域，（2）编码啤酒酵母 α 交配因子（AMF）前原序列和编码HSA的核苷酸序列，（3）在甲基营养型酵母中有功能活性的转录终止子的，以及（4）至少一个可选择标志基因和细菌复制原点组成的表达盒。

本发明的再一目的是提供用上述携带HSA编码序列的重组表达载体转化的甲基营养型酵母细胞。

根据本发明这一目的的一个优选实施方案, 其中所说的甲基营养型酵母是巴斯德毕赤酵母细胞。

根据本发明这一目的的一个优选方案, 其中所说的酵母细胞可以是具有 Mut⁺或 Mut⁻表型的。

本发明的再一个目的是提供以重组 DNA 技术生产人血清白蛋白的方法，该方法包括在适于表达所需的蛋白质的条件下培养按上述方法制备的包含编码人血清白蛋白之核苷酸序列的酵母宿主细胞，然后从培养物中回收并纯化所需的人血清白蛋白。

本发明的再一个目的是提供按上述方法生产的,基本上具有 SEQ ID No: 2 所示氨基酸序列的人血清白蛋白。

图 1A 显示合成的编码全长度人血清白蛋白的核苷酸序列及由之推断的人血清白蛋白的氨基酸序列:

图 1B 显示图 1A 中所示全长度人血清白蛋白核苷酸编码序列中的单一核酸内切酶酶切位点:

图 2 显示编码人血清白蛋白的核苷酸序列的合成策略:

图 3 显示 HSA-I 片段的核苷序列(上行)及由之推断的氨基酸序列(下行);

图 4 显示重组质粒 pBKS/HSA-I 的构建:

图 5 显示 HSA-II 片段的核苷酸序列(上行)及由此推断的氨基酸序列(下行);

图 6 显示重组质粒 pBKS/HSA-II 的构建:

图 7 显示 HSA-III 片段的核苷酸序列 (上行) 及由之推断的氨基酸序列 (下行);

图 8 显示重组质粒 pBKS/HSA-III 的构建;

图 9 显示重组表达载体 pBKS/HSA 的构建;

图 10 显示重组表达载体 pPIC9/HSA 的构建;

图 11 显示 HSA 核苷酸编码序列与表达载体 pPIC9 连接的连接区序列;

图 12 显示 HSA-I、HSA-II、HSA-III 片段及全长度 HSA DNA 片段的 1 % 琼脂糖凝胶电泳图谱;

图 13 显示合成的 HSA 基因在巴斯德毕赤酵母菌株 GS115 中表达的时间过程, 其中泳道 1 - 8 分别显示在培养毕赤酵母转化体细胞的培养基中加入甲醇诱导 HSA 表达后 0、24、24、48、72、96、120 和 148 小时收集培养物上清并在 10 % 聚丙烯酰胺凝胶上进行分析的结果;

图 14 显示对 HSA 表达产物所作的免疫印迹分析的结果, 其中各泳道加样材料分别为 A: 天然 HSA; B: 不携带重组质粒的 GS115 酵母细胞的培养物上清; C: 用 pPIC9/HSA 转化的 GS115 酵母细胞的培养物上清。

就以 DNA 重组技术生产异源蛋白质, 特别是对纯度要求极高且分子中富含巯基的人血清白蛋白来说, 酵母的最突出的优点之一是其能够向培养基中分泌异源蛋白质产物。因此, 使用酵母表达系统能够以较高的起始纯度得到被分泌的产物, 从而可简化纯化步骤。特别是对于富含巯基的 HSA 来说, 建立能够向培养基中分泌表达产物的酵母表达系统的另一个具有说服力的理由在于 HSA 产物更易于通过二硫键形成来维持其准确的四维结构。细胞和细胞外介质的分泌途径是能够支持二硫键形成的氧化环境 (Smith, et al., Science 229:1219(1985))。相反, 胞浆则是一个不能形成二硫键的还原环境。细胞破碎后, 如果二硫键形成过快, 可导致随机二硫键的生成。因此, 最好是通过分泌途径生产富含二硫键的蛋白质, 如 HSA。

为了克服某些与啤酒酵母有关的问题, 例如质粒选择标志的丢失以及在高密度发酵条件下的质粒分配、拷贝数和稳定性等问题, 进而建立了基于甲基营养酵母, 如巴斯德毕赤酵母的酵母表达系统。这一系统的一个关键性特征是其利用衍生于巴斯德毕赤酵母的甲醇调节的醇氧化酶-1 (AOX1) 基因作为启动子, 驱动外源基因的表达 (参见欧洲专利申请 183071)。该表达系统的另一个特征是可将表达盒稳定的整合到细胞染色体中, 从而显著地降低载体丢失的机会。

在本发明之前, 不仅已使用巴斯德毕赤酵母成功的生产了例如乙型肝炎表面抗原 (Cregg et al., Bio/Technology 5:479(1987))、溶菌酶和转化酶 (Digan et al. Developments in Industrial Microbiology 29:59(1988); Tschopp et al., Bio/Technology 5:1305(1987)), 以及表皮生长因子 (国际专利 WO90/10697), 而且也已在甲基营养酵母, 特别是巴斯德毕赤酵母中成功地表达了人血清白蛋白 (欧洲专利申请 0344459)。现有技术中所使用的 HSA 基因序列基本上是基于 Lawn 等人 and Dugaiczky 等人前已分离并测序的

HSA 基因序列。然而，本发明人的早期研究发现，虽然 HSA 基因更适于在真核生物体包括在酵母中表达，但考虑到人细胞与低等真核生物体如酵母细胞在结构和蛋白质的生物合成机制例如密码子使用频率、基因转录的启动与转录效率、RNA 的剪切与翻译效率以及翻译产物的分泌等方面存在很大差异，因而有必要对天然 HSA 基因进行某些修饰，使之更适于在酵母系统，特别是毕赤酵母细胞中稳定地存在，并更为有效地转录和转译。

另外，一些研究表明，不同生物体在密码子使用频率上存在很大差异，从而直接影响转译效率和表达水平。如果对异源基因进行适当的改造，使待表达基因的密码子更适合于所选择的表达系统的密码子使用频率，必将有利于提高外源基因的总表达效率。Kerlak 等人 (Bio/Technology 8:939-42, (1990)) 按照双子叶植物的密码子使用频率修饰苏云金芽孢杆菌 (B.t) β 内毒素基因，并将该基因转化到双子叶植物如棉花细胞中表达，结果导致该修饰的 B.t 内毒素基因在棉花中的表达水平比天然 B.t 内毒素基因的表达水平提高 50-100 倍。相反，Hoekeman 等人将磷酸甘油激酶基因克隆到酵母表达载体中，并将其 5' 端的部分酵母偏性密码子用稀有密码子取代，结果使之在酵母细胞内的表达效率降低了大约 10 倍。

本发明人充分考虑了密码子使用频率对基因表达的影响和天然 HSA 编码序列的可能存在的不利因素，全面修饰并重新合成了 HSA 基因。为此，首先按照在密码子使用频率上基本上与毕赤酵母相同的啤酒酵母的密码子使用频率，重新设计并合成了全长度人血清白蛋白核苷酸编码序列。与天然 HSA 基因序列相比较，本发明合成的基因序列中有 155 个碱基被其它碱基所取代，并因此而导致 141 个密码子的改变。本发明合成的 HSA 基因的密码子使用频率如下列表 1 所示。

表 1：合成的 HSA 基因的密码子使用频率

密码子	氨基酸	使用频率	密码子	氨基酸	使用频率	密码子	氨基酸	使用频率	密码子	氨基酸	使用频率
TTT	phe	19	TCT	ser	8	TAT	tyr	9	TGT	cys	22
TTC	phe	13	TCC	ser	5	TAC	tyr	9	TGC	cys	13
TTA	leu	12	TCA	ser	4	TAA	och	1	TGA	opa	-
TTG	leu	18	TCG	ser	1	TAG	amb	1	TGG	trp	1
CTT	leu	10	CCT	pro	8	CAT	his	11	CGT	arg	3
CTC	leu	1	CCC	pro	3	CAC	his	5	CGC	arg	1
CTA	leu	10	CCA	pro	11	CAA	gln	14	CGA	arg	2
CTG	leu	11	CCG	pro	2	CAG	gln	6	CGG	arg	1
ATT	ile	4	ACT	thr	11	AAT	asn	9	AGT	ser	5
ATC	ile	4	ACC	thr	7	AAC	asn	8	AGC	ser	2
ATA	ile	1	ACA	thr	7	AAA	lys	32	AGA	arg	14
ATG	met	6	ACG	thr	3	AAG	lys	28	AGG	arg	5
GTT	val	15	GCT	ala	27	GAT	asp	26	GGT	gly	9
GTC	val	9	GCC	ala	15	GAC	asp	11	GGC	gly	2
GTA	val	6	GCA	ala	16	GAA	glu	42	GGA	gly	3
GTG	val	11	GCG	ala	4	GAG	glu	21	GGG	gly	-

为了便于将本发明的合成的 HSA 基因连接到各不同的克隆载体或表达载体上, 我们在其 5' 和 3' 端分别加入两个单一酶切位点, 即 5' 端加入 BamHI 和 XhoI 酶切位点, 3' 端加入 EcoRI 和 ClaI 酶切位点, 其中 BamHI 和 ClaI 切点用于连接克隆载体, 而 XhoI 和 EcoRI 则用于连接表达载体. 另外, 为了确保酶切位点的有效利用, 我们分别在合成的基因序列的 5' 端和 3' 端加入 3 个保护性碱基 (分别是 GGC 和 AGG). 再者, 为了便于将合成的 HSA 基因连接到信号肽的 C 末端上, 还在该基因的 N 末端附加了氨基酸序列作为信号肽识别和切割位点. 作为 HSA 基因的必要调控元件, 其 C 末端带有一个重复的终止密码子序列, 借以确保 mRNA 转译过程的正确终止. 上述这些修饰导致本发明合成的 HSA 基因序列总长度约为 1797bp.

另一方面, 为了进一步提高本发明合成的 HSA 基因在甲基营养型酵母, 特别是巴斯德毕赤酵母中的表达效率, 还对基因之核苷酸序列的下述几个可能对基因转录和/或转译效率带来不利影响的序列或其他部分进行了必要的修饰, 改动, 置换或删除.

1、内含子剪接序列: 真核生物具有对基因转录产物进行剪接加工以形成不带内含子的 mRNA 的能力. 现已知道, 内含子的剪接是由特殊序列介导的, 如果 DNA 序列中存在相应的剪接位点, 则这样的序列即可在基因转录过程中被作为内含子从序列中切掉, 以致产生不完全的转录产物. 因此, 在基因的设计和合成中, 有必要检查并排除这样的碱基序列.

2、转录终止序列: 如果在全长度基因序列内部非转录终止部位存在额外的转录终止序列, 便有可能导致转录的提前或错误终止, 以致产生不完整的 mRNA. 在酵母表达系统中, 已知与转录终止相关的序列包括: (1) ATTATTTTAT; (2) TTTTTATA; (3) TAG...(富含 T)...TA(T)GT...(富含 AT)...TTT. 因此, 在设计并合成 HSA 基因时, 必须排除这些序列, 特别富含 AT 的序列.

3、长的重复序列: 重复序列, 特别是长的重复序列 (包括正向和反向重复序列) 有可能形成强的二级结构, 从而不利于基因转录和转译的顺利进行, 而且有可能给全长基因的拼接造成困难. 因此, 当序列中出现七个以上碱基的重复序列时, 需在不改变其所编码的氨基酸的前提下通过改变所牵连的密码子而排除之.

4、内部启动子序列和核糖体结合位点: 基因序列的内部可能含有不期望存在的驱动基因内部转录过程的启动子序列, 这些序列有可能干扰全长度基因的正常转录, 因此必须予以除去. 另外, 如果基因内部存在多余的核糖体结合位点, 也有可能启动基因内部的转译过程, 以致产生不完整的蛋白质表达产物. 因此, 在本发明的合成 HSA 基因序列中排除了 TATA 框

启动子序列和 GGAGG 核糖体结合位点。

此外，为了便于拼接预合成的多核苷酸片段，并便于将串联拼接的全长度基因序列在适当的酶切位点处与其他基因转录和/或转译调控元件拼接或方便的插入到选择的表达载体中，还有必要在基因序列中适当的核苷酸位置上加入若干个酶切位点或接头序列。因此，除上文提到的 5'端的 BamHI 和 XhoI 酶切位点，以及 3'端 EcoRI 和 ClaI 酶切位点外，还在核苷酸位置 509 和 1109 处分别加入了 HindIII 和 PstI 位点（图 1B）。

因此，本发明的第一个目的是提供按上述原则合成的编码人血清白蛋白的核苷酸序列。为了得到如图 1A 中所示的编码人血清白蛋白的全长度核苷酸序列，合成了 3 个核苷酸长度分别为 612bp、562bp 和 705bp 并分别定名为 HSA-I、HSA-II 和 HSA-III 的片段。其中每个片段均由 8 条较短的寡核苷酸连接而成，并且各片段的两端均分别带有用于在适当的克隆载体中直接克隆并克隆的 BamHI 和 ClaI 酶切位点。

可以按照本领域技术人员已知的常规方法分别化学合成相对短的寡核苷酸序列的互补链，并经变性后退火得到相应的双链 DNA 分子。但从经济和方便的角度考虑，最好采用本发明人改进的基于 PCR 技术的 DNA 序列合成方法合成本发明的 HSA 基因序列（参见实施例 1）。按照这种方法，两条互补链之间只保留 < 20 个碱基对的重叠互补区，即足可以一条模板链为基础，合成长达 120 个碱基的双股寡核苷酸序列。这样，经一次的拼接即可得到正确连接的长达大约 900bp 的 DNA 片段。

由于在利用 Oligo4.0 计算机软件设计各寡核苷酸链时，充分分析了相邻两条寡核苷酸链上的重叠区序列，从而有效的避免了所说的重叠区内发夹结构和强二级结构的形成，增加了重叠区序列的特异性的分子内部稳定性。一般说来，可通过改变重叠区内的个别碱基或造成总体重叠区的位移达到这一目的。

图 2 图解显示了本发明的 HSA 核苷酸编码序列的合成策略。简单的说，首先合成 3 个分别定名为 HSA-I、HSA-II 和 HSA-III 的长度分别为 612bp、362bp、和 705bp 的 DNA 片段（参见图 3、5、和 7）。这些片段各自均由 8 个寡核苷酸序列组成。然后将这三个片段分别克隆到携带 M13 噬菌体 DNA 复制原点、氨苄青霉素抗性基因及多克隆位点，并且在多克隆位点两侧分别含有 T3 及 T7 噬菌体启动子的 pBKS 质粒载体中，得到分别定名为 pBKS/HSA-I、pBKS/HSA-II 和 pBKS/HSA-III 的 3 个重组质粒（参见图 4、6、和 8）。从这些质粒中提取 DNA 并分别测定核苷酸序列后，按上述原则以适当方法纠正个别碱基的错误。待确定各个片段的碱基序列后，利用各片段两端预先设定的 HindIII 和 PstI 酶切位点将它们连接到利用 BamHI 和 ClaI 切割后线性化的质粒 pPKS 中，在 T₄DNA 酶作用下重新环化

后得到重组质粒 pBKS/HSA (图 9)。转化到大肠杆菌 DH5 α 中进行多次传代后,从增殖的细菌细胞中提取 DNA,并以双脱氧链终止法测定全长度 HSA 基因的核苷序列,结果表明本发明合成的 HSA 可稳定保留在继代培养的大肠杆菌细胞中并且没有发生任何序列突变和/或缺失。

用于表达本发明合成的 HSA 基因的表达载体可以是携带适于在酵母中表达外源基因之表达调控元件,并可在酵母特别是甲基营养型酵母中以高拷贝数插入的任何载体,例如 pPIC9(Invitrogen)、pPIC3(Invitrogen)、pHIL-D2(Invitrogen)、yYJ32(NRRL B-15891)、pJDB207(Amershan)、pJDB219(Amershan)、PG4.0(NRRL B-15868)、为本发明的目的,其中优选的是 pPIC9。

例如可用 EcoRI 和 XhoI 双酶消化,将 pPIC9 质粒切成线性,并与用同样双酶切后从 pBKS/HSA 中得到的 HSA 基因片段连接,然后用所得连接混合物转化适当的大肠杆菌细胞,在 LB 培养基平板上 37 $^{\circ}\text{C}$ 培养并筛选被转化的重组体细胞后,从中提取质粒 DNA 并根据片段大小鉴定重组质粒。

因此,本发明进一步涉及至少含有一个表达盒的重组表达载体,其中所说表达盒在转录的读框方向上依次含有下列 DNA 序列:

- (i) 甲基营养型酵母之甲醇反应性基因的启动子区域;
- (ii) 编码啤酒酵母 α 交配因子 (AMF) 前原序列和编码 HSA 多肽序列的 DNA 序列,以及
- (iii) 在甲基营养酵母中有功能活性的转录终止子。

可将此 DNA 片段作为侧翼接有与靶基因有足够同源性之 DNA 序列的线性片段,转化到甲基营养型酵母如巴斯德毕赤酵母中,以实现所说的 DNA 片段在酵母染色体中的整合。在这种情况下,整合是通过重组表达载体中的同源序列在酵母靶基因组的适当位点处与酵母靶基因组发生同源重组而实现的。或者所说的 DNA 片段可以是环形质粒的一部分,可将其切成线性以有利于整合,并且可以在宿主和质粒序列间的同源性位点处加入该片段进而实现其在宿主基因组中的整合。载体的线性化将有利于外源基因的整合,并且常常在线性化位点处插入宿主染色体。

用于转化甲基营养型酵母细胞的表达盒,除含有甲基营养型酵母甲醇反应性启动子和 HSA 编码 DNA 序列 (HSA 基因) 外,还含有编码读码内啤酒酵母前原序列切割识别位点 (即 Glu-Lys-Arg 编码序列),以及可在甲基营养型酵母中发挥功能的转录终止子序列。

用于驱动 HSA 基因表达的优选启动子区衍生于巴斯德毕赤酵母的甲醇诱导或调节的醇氧化酶(AOX)基因。已知巴斯德毕赤酵母含有醇氧化酶 I (AOX1) 和醇氧化酶 II (AOXII) 基因。两个 AOX 基因的编码部分在

DNA 和推断的氨基酸序列水平上有很大同源性, 并具有共同的限制性酶切位点. 由两基因启动表达的蛋白质有相似的酶学性质, 但其中 AOX1 基因的启动子更为有效. 因此, 为表达 HSA 目的, 优选的是使用 AOX1(Ellis et al., Mol. Cell.Biol. 5:1111(1985)).

酵母 α 交配因子或称 α 因子是由“ α ”交配型的细胞分泌的含 13 个氨基酸的肽. 其作用于相对“ a ”交配型的细胞, 以促进两种类型的细胞之间的有效接合, 从而形成“ $\alpha - a$ ”二倍体细胞 (Thorer et al., The Molecular Biology of The Yeast *Saccharomyces*, Cold Spring Harbor Laboratory, Cold Spring Harbor, 143(1981)). AMF 前原序列是包含在 AMF 前体中的前导序列并包括有蛋白质水解加工和分泌所需的 Glu-Lys-Arg 编码序列. AMF 前原序列 (包括 Leu - Lys - Arg 编码序列) 是一个 258bp 的片段.

根据本发明在甲基营养型酵母中表达 HSA 的表达盒还含有转录终止子, 其具有一个编码多聚腺苷酸信号和聚腺苷酸化位点的亚片段, 和/或为从启动子起始转录提供转录终止信号的亚片段.

上述用于本发明的 HSA 表达和分泌的功能性调控序列 (非结构基因调控元件) 一般均已包括在本发明为表达 HSA 目的所选用的表达载体 pPIC9 (Invitrogen) 中, 并且其中转录终止子与启动子的来源可以是相同或不同的.

根据本发明的表达载体, 其中还可包括可在甲基营养型酵母如巴斯德毕赤酵母中发挥功能的选择标志基因. 为此, 所说的选择标志基因可以是任何赋予甲基营养型酵母如巴斯德酵母以某种表型, 从而得以鉴定它们并使之获得未被转化的细胞所不具有的生长能力的基因. 例如, 适当的选择标志基因包括由营养缺陷型突变体巴斯德毕赤酵母宿主菌株和弥补宿主缺陷的野生型生物合成基因组成的选择标志系统. 例如当转化 $HIS4^-$ 巴斯德毕赤酵母菌株时, 可利用啤酒酵母或巴斯德毕赤酵母 $HIS4$ 基因.

如果用含有处于巴斯德毕赤酵母基因启动子调节下之 HSA 基因和加工与分泌所需之 AMF 序列的线性 DNA 片段转化酵母宿主, 则上述表达盒将通过已知的一步骤或二步骤置换技术 (如参见 Rothstein, Methods Enzymol. 101:202(1983); Sherer and Davis, Proc.Natl.Acad.Sci., USA, 76:4951(1979)) 被整合到宿主基因组中, 可借助与靶基因有足够同源性的侧翼 DNA 序列将线性 DNA 片段导向特定的座位, 以实现 DNA 片段在其中的整合.

如果含有本发明合成的 HSA 编码序列的 DNA 片段包含在表达载体如环形质粒内, 则可在基因组的相同或不同座位上整合一个或多个质粒拷贝. 在含有本发明合成的 HSA 编码序列的 DNA 片段被整合前或整合后, 其中表达盒的各部分是彼此可操作地连接的. 对于启动子、编码啤酒酵母 AMF 加工位点的 DNA 序列及转录终止子来说, 编码 HSA 的 DNA 序列是

位于其功能性位置和方向上的。因此，HSA 编码序列可在启动子区的控制下被转录成能够在转译后产生所需的 HSA 的转录物。因为表达载体内存在 AMF 前原序列，所以可从培养基中分离得到作为被分泌的整体的 HSA 产物。由于含有本发明合成的 HSA 基因序列的 DNA 片段包括允许其在细菌特别是大肠杆菌中复制的序列，所以可在细菌中大量产生该 DNA 片段。

根据本发明这一目的的一个优选实施方案，可将按本发明上述方法合成的编码 HSA 的 DNA 序列可操作地连接到可从市场上购得的，携带适于在甲基营养型酵母中表达和分泌外源蛋白质的功能性调控序列（非结构基因调控元件）的表达载体，例如可购自 Invitrogen 公司的 pPIC9 中。该质粒含有 5'AOX1 启动子片段、 α -因子分泌信号、多克隆位点区、3'AOX1 转录终止片段、HIS4 ORF、3'AOX1 片段、COLE1 复制原点及氨苄青霉素抗性基因。转化甲基营养型酵母例如巴斯德毕赤酵母的方法，以及用于培养其基因组中含有编码外源蛋白质之基因的甲基营养酵母，例如巴斯德毕赤酵母的方法是本技术领域已知的。根据本发明，可用 Cregg 等人（Mol.Cell.Biol. 55:3376(1985)）、所述的原生质体技术或全细胞氯化锂酵母转化系统（ITO, et al, Agric. Biol.Chem. 48:341(1984)）以及电穿孔技术将表达载体转化到甲基营养型酵母细胞中。在使用单转录单位的情况下，较好选用电穿孔转化技术。

用 Southern 印迹分析法（Sambrook et al.,Molecular Cloning,2nd ed.,Chapter9, Cold Spring Harbor laboratory Press,Cold Spring,NY,USA,1989）分析 DNA 整合位点，并用 Northern 印迹法（Sambrook 等人,上述文献,第 7 章）分析甲醇反应性 HSA 基因表达,以进一步鉴定阳性转化体细胞。

本发明进一步提供利用甲基营养型酵母表达系统生产重组 HSA 的方法，该方法包括在适于表达所需的蛋白质的条件下培养按上述方法制备的包含编码 HSA 之核苷酸序列的酵母宿主细胞，然后从培养物中回收并纯化所说的 HSA。为此，在发酵器中培养带有预期的表型和基因型的被转化的菌株。为了由甲基营养型酵母如巴斯德毕赤酵母大规模生产基于重组 DNA 技术的产品，一般可采用三阶段高密度分批发酵系统。第一阶段即生长阶段是在以过量甘油作为碳源的培养基中培养宿主细胞。在这种碳源上生长时，酵母宿主中的异源基因的表达受到限制，从而使细胞在没有产生异源蛋白质的情况下增殖。然后是较短时间的甘油限制生长阶段。在甘油限制生长阶段以后，加入甲醇诱导所需的异源蛋白质的表达，即为生产阶段（第三阶段）。图 13 显示了本发明合成的 HSA 基因在巴斯德毕赤酵母中表达的时间过程。

根据本发明的一个优选实施方案，用于生产 HSA 的异源蛋白质表达系统利用了衍生于巴斯德毕赤酵母的甲醇调节之 AOX1 基因的启动子。该启动子可有效地启动基因的表达并严格接受外源营养因子的调节。所说的基因也可以是转录终止子的来源。本发明优选的表达盒包括可操作地彼此连接的巴斯德毕赤酵母 AOX1 启动子、编码啤酒酵母 AMF 前原序列（或称 α 因子）的 DNA、按本文前述原则全合成的编码成熟 HSA 的 DNA 序列，以及衍生于巴斯德毕赤酵母 AOX1 基因的转录终止子。在一个 DNA 片段上最好含有两个或多个头尾连接的这样的表达盒，以在单一连续 DNA 片段上产生多表达盒。

如前所述，为了提高 HSA 基因序列在巴斯德毕赤酵母中的表达效率和表达水平，本发明人充分考虑了天然存在的 HSA 编码序列的某些结构特征及其可能存在的对表达效率的不利影响，全面修饰并重新合成了所说的 HSA 基因。这些修饰包括用酵母偏性密码子取代天然 HSA 基因中的所有非酵母偏性密码子、加入适当的便于片段连接与重组的核酸内切酶位点、粘端加入必要的保护性碱基和信号肽识别序列与切割位点，以及附加的终止密码子。另一方面，本发明还针对 HSA 基因序列中的内含子剪接序列、转录终止序列中长的重复序列、内部启动子序列及核糖体结合位点等可能对基因转录和/或转译效率带来不利影响的部分进行了必要的碱基修饰、改动、和置换。

可将本发明的按上述原则合成的 HSA 基因序列插入适当的表达载体如 pPIC9 中，并用所得重组载体转化适当的大肠杆菌细胞，从被转化的大肠杆菌中分离质粒 DNA 并进行序列分析。分析结果显示本发明合成的 HSA 被正确地连接到表达载体所携带的 AOX1 基因的 5' 启动子序列和 3' 转录终止子序列之间，并以 5' 端直接连接到包括 AMF 前原序列（ α 因子编码序列）之蛋白水解加工位点 Glu-Lys-Arg 的 3' 编码序列上（参见图 11）。

上述合成基因对天然 HSA 编码序列的修饰和改动以及其在适当表达载体中的正确连接，充分保证了基因表达的有效启动、顺利转录和转录终止，同时也确保了转录产物之转译过程的顺利进行。

可用上述表达盒转化的优选的宿主细胞是具有可用存在于转化 DNA 片段上的标志基因弥补的至少一种突变的巴斯德毕赤酵母细胞，最好是利用 HIS4⁻（GS115）或 ARG4⁻（GS190）营养缺陷型突变体巴斯德毕赤酵母菌株。将含有一个或多个表达盒的 DNA 片段插入到含有弥补宿主之缺陷的标志基因的质粒中。本发明优选的表达载体是质粒 PAO815 或 pPIC9，特别是 pPIC9。

为了建立巴斯德毕赤酵母的 Mut^- 表达菌株, 可用一步骤基因置换技术将含有表达盒的转化 DNA 整合到宿主基因组中, 用适当的核酸内切酶消化表达载体以产生两末端同源于 AOX1 座位的线性 DNA 片段(借助侧翼同源序列)。经这样的基因置换即得到 Mut^+ 菌株, 这里所说的 Mut 是指甲醇利用表型。在 Mut^+ 菌株中, AOX 基因被表达盒所取代, 因而降低了菌株利用甲醇的能力。在这种情况下, 酵母菌株便通过 AOX2 基因产物的表达来维持其在甲醇上的缓慢生长。可根据补偿基因的存在来鉴定其中表达盒已通过位点特异性重组而整合到 AOX1 座位中的转化株。

为了建立表达 HSA 的 Mut^+ 菌株, 最好用包括表达盒的环形质粒转化宿主细胞, 使包括一个或多个表达盒的 DNA 片段整合到宿主基因组中。在这种情况下, 整合是通过在与转化载体上的一个或多个序列有同源性的座位上加入来实现的。为了鉴定阳性转化株, 可以用 Southern 印迹法分析 DNA 整合的位点, 用 Northern 印迹法分析甲醇反应性 HSA 基因的表达, 并通过检测培养基中的 HSA 产物来证实 HSA 蛋白质的分泌表达。可用 Southern 印迹法鉴定在预期的位点整合了一个或多个表达盒的巴斯德毕赤酵母菌株。可用 Northern 印迹法和蛋白质产物分析法鉴定提高了 HSA 分泌产率的酵母菌株(参见图 14)。

在发酵罐中, 例如在改良的 YNB 或 YPD 液体培养基中, 于大约 30 °C 的温度下培养已证明具有预期的基因型和表型的甲基营养型酵母转化株。可使用典型的三阶段生产方法发酵生产 HSA, 首先使细胞生长在抑制性碳源, 最好是在过量甘油上。这个阶段是在没有表达的情况下产生细胞群体。然后使细胞在限制甘油供应条件下生长。当甘油被消耗后, 向培养器内加入单一甲醇或有限量甘油 + 甲醇, 以导致由甲醇反应性启动子驱动的 HSA 基因的表达。为了以高细胞密度发酵并提高产物的表达水平, 可在此阶段以补料分批培养的方法加入少量的高浓度蔗糖和葡萄糖以避免高浓度底物对细胞生长的抑制, 必要时亦可以向培养基中加入脂肪酸如有 6-18 个碳原子的脂肪酸来改善 HSA 产物的产率。

分离和收获 HSA 的方法可包括加热处理上述培养物上清以灭活蛋白酶, 并使用选自阳离子交换剂、疏水载体及活性碳的至少一种材料将 HSA 与着色成份分离开, 以抑制产物的颜色生成。可以在培养期间向培养基中加适当的胺如甲胺、乙胺或丙胺以抑制 HSA 的颜色。

可使用选自加热处理、超滤、亲和层析、阳离子交换层析、疏水层析、阴离子交换层析、盐析、和螯合树脂处理的至少五种组合方法纯化得到高纯度 HSA。为了适当的简化纯化步骤, 尽可能地在保证 HSA 产物纯度的前提下降低生产成本, 较好选用依次加热处理、超滤、疏水层析、和阴离

子交换层析或高压液相层析的连续纯化方法。按本发明方法生产的 HSA 产率一般约为 3.5g/升培养液，产物纯度基本可达到 99.99%。

下列实施例旨在进一步举例描述本发明，但这些实施例均不以任何方式构成对本发明待批权利要求之范围的限制。

实施例 1：HSA 基因全长序列的合成与克隆

本实施例描述编码 HSA 的全长度核苷酸序列的合成。鉴于全长 HSA 基因的核苷酸序列相对较长，一次合成可能遇到操作上的困难，因此将全长基因序列分为三个较大的核苷酸片段分别合成。这三个片段分别称为 HSA-I (612bp)，HSA-II (562bp)，和 HSA-III (705bp)，并且它们每个都由 8 个寡核苷酸片段拼接而成。

1. HSA-I 片段的合成与克隆

(1)为了合成长度为 612bp 的包括 HSA 基因之 5'侧的片段 HSA-I，首先使用 ABI394 型 DNA 自动合成仪 (Applied Biosystem Inc.)按仪器制造商推荐的步骤在 1000Å固相载体上合成有下示序列的 8 个寡核苷酸链：

OA-1
GGCGGATCCCTCGAGAAAAGAGATGCACACAAGAGTGAGGTTGCTCATCGGTTT
AAAGATTTGGGTGAGGAAAATTTCAAAGCCTTAGTGTTAAT
OA-2
GCGAACTCAGTGACCTCGTTAACCAACTTTACATGATCCTCAAATGGACATTGCT
GAAGATATTGGGCAAAGCAATTAACACTAAGGCTTTG
OA-3
CGAGGTCAGTTCGCTAAAACCTTGTGTTGCAGACGAATCTGCTGAAAATTG
TGATAAGTCTCTGCATACCCTATTTGGTGACAAATTGTGC
OA-4
GCATTCGTTTCTGCCAGGTTCTTGCTTTGCACAGCAGTCGGCCATTTACCGTAA
GTTTCTCGAAGGGTAGCCGACTGTGCACAATTTGTCACCAA
OA-5
CCTGGCAGAAACGAATGCTTTTTGCAGCACAAACATGACAACCCTAATCTTCCA
CGTTTAGTGAGACCTGAAGTTGATGTGATG
OA-6
TATGGGTGCCTTCTTGCAATTTTCATATAAGTACTTCTTCAAAAAAGTTTCTTCATT
ATCATGGAAGGCTGTACACATCACATTCAACTTCAGG
OA-7
GCAAGAAGGCACCCATACTTCTATGCCCCGGAGCTACTGTTCTTTGCTAAACGTT
ATAAAGCTGCATTACAGAATGTTGCCAAGC
OA-8
CCTATCGATGGCAGAAGAAGCTTTACCCCTCATCTCTTAATTCGTCTAGCTTTGGC
AACAGACAAGCTGCCTTATCGGCAGCTTGGCAACATTCTGTG

将分别合成的寡核苷酸 (OA-1 至 OA-8) 从固相载体柱上洗脱并用含 7M 尿素的 10%聚丙烯酰胺凝胶电泳纯化后，分别回收之。

按照图 2 所示的合成策略, 使用 PCR 反应介导的 DNA 合成技术, 由上述 8 个寡核苷酸经三步骤 PCR 扩增得到 HSA-I 片段。具体地说, 首先在四个 PCR 反应试管中分别加入 OA-1 和 OA-2 (各 $10\mu\text{l}$)、OA-3 和 OA-4 (各 $10\mu\text{l}$)、OA-5 和 OA-6 (各 $10\mu\text{l}$) 以及 OA-7 和 OA-8 (各 $10\mu\text{l}$)。然后向总反应体积为 $50\mu\text{l}$ 的各管内分别加入约 $50\mu\text{l}$ $1\times$ PCR 缓冲液, $200\mu\text{M}$ dNTP (各 $50\mu\text{M}$), 1.5mM MgCl_2 和 2.5 单位 Taq DNA 聚合酶。在 94°C , 10 秒; 55°C , 30 秒; 72°C , 1 分钟的 PCR 反应条件下, 经 15 次 PCR 循环得到 8 条寡核苷酸链两两连接的 4 个双链 DNA 片段: 片段 A、B、C、和 D。然后将片段 A 和 B、C 和 D 分别混合在两个反应容器内, 再次按上述条件进行 PCR 反应, 分别得到进一步融合并扩增的片段 E 和 F。最后, 使用寡核苷酸 OA-1 和 OA-8 作为两末端引物, 再次对如上得到的片段 E 和 F 的混合物进行 PCR 反应, 从而得到代表 HSA 基因的 5' 端的片段 HSA-I。用 1.5% 琼脂糖凝胶分离并纯化该片段后, 置 -70°C 下保存备用。

(2) 将携带质粒 pBKS 的大肠杆菌 DH5 α 菌株接种于 50ml LB 培养基中, 37°C 振荡培养 15 小时。培养后离心 (5000rpm , 5 分钟) 收集细胞, 并以碱裂解法制备用于克隆合成的 HSA 片段的质粒 pBKS。为此, 首先将细胞沉淀物悬浮于 5ml 溶液 I (50mM 蔗糖, 25mM Tris, 10mM EDTA, $\text{pH}8.0$) 中, 然后向悬液中加入 10ml 新鲜的溶液 II (0.2NaOH , 1% SDS), 混匀并置冰水浴中放置 10 分钟后再加入 7.5ml 溶液 III (3M KAc, $\text{pH}4.8$), 再次混匀于冰浴中放置 15 分钟, 离心 (12000rpm , 10 分钟) 收集沉淀物。用 70% 乙醇洗细胞沉淀物并重新悬浮于 1ml TE 缓冲液中。向所得溶液内加入 RNA 酶 A (Promega) ($200\mu\text{g/ml}$) 并于 37°C 保温 30 分钟。然后依次用苯酚: 酚: 氯仿 (1:1) 和氯仿: 异戊醇 (24:1) 抽提, 并向提取液内加入 1/10 体积的 3M NaAc ($\text{pH}4.8$) 和两倍体积的无水乙醇以沉淀之。离心后用 70% 乙醇洗涤沉淀物并重新溶解于 0.1ml TE 缓冲液中, 从而得到 pBKS 质粒 DNA。最后用限制性核酸内切酶 BamHI 和 ClaI (各 2 单位) 消化该质粒 DNA, 用 1% 琼脂糖电泳分离并回收所得的 3.0Kb 的酶切产物, 得到纯化的线性化 pBKS 载体。

(3) 用 BamHI 和 ClaI 双酶 (各 1 单位) 酶切如前制备的 HSA-I 片段 ($4\mu\text{g}$), 将此 HSA-I 片段的酶切产物 ($1\mu\text{g}$) 与如上同样酶切的 pBKS 质粒 (约 $4\mu\text{g}$) 混合, 在 T4 DNA 连接酶 (3 单位) 存在下, 于 16°C 水浴中反应 17 小时。反应完成后按常规方法用所得连接混合物转化感受态大肠杆菌 DH5 α 菌株的细胞悬液 (OD_{600} 约为 0.6)。转化反应中使用 pBKS 空质粒作为阴性对照。将细菌细胞悬液铺敷在添加氨苄青霉素 ($100\mu\text{g/ml}$) 的固体 LB 平板上, 37°C 保温过夜后, 筛选出被转化的阳性菌落, 并

制备质粒 DNA。用 BamHI 和 ClaI 酶切该 DNA，根据所得 DNA 片段的大小初步鉴定以正确方向连接有 HSA-I 片段的重组质粒 pBKS/HSA-I。

(3) 在添加氨苄青霉素 (100 μ g/ml) 的固体培养基上培养经初步鉴定的携带重组质粒 pBKS/HSA-I 的阳性菌株，并按上述方法从中制备 DNA 并以此 DNA 作为下述测序反应的模板。按已知方法对模板 DNA (2 μ g) 进行变性处理，然后向其中加入测序引物和退火缓冲液 (各 2 μ l) 以完成退火反应，并按已知的 Sanger 双脱氧链终止法 (Sambrook et al., Molecular Cloning, 2nd Edition Chapter 13, Cold Spring Harbor laboratory Press, Cold Spring, NY, USA, 1989) 检测变性的双链 DNA 模板的核苷酸序列，测序结果显示所有五个克隆均包括有数目不等的碱基缺失和置换。为此利用 HSA-I 系列中的适当位置上存在的酶切位点，切出没有碱基错误的小片段并与其他克隆中带有正确序列的小片段重新连接后得到去除了错误碱基的 HSA-I 片段。在纠正 HSA-I 片段的碱基错误中，选择使用了两个内部单一酶切位点即核苷酸位置 282 处的 HaeIII 和核苷酸位置 431 处的 ScaI 位点。

2. HSA-II 片段的合成与克隆

按照上文 1(1) 中所述合成 HSA-I 片段的相似方法，首先合成具有下示 8 个寡核苷酸：

OA-9

GGCGGATCCGAGGGTAAAGCTTCTTCTGCCAAACAAAGATTGAAGTGCGCTAGT
CTACAGAAAT

OA-10

AAACTCGGCTTTGGGAAATCTCTGGCTCAGGCGAGCTACAGCCCATGCCTTAAA
AGCTCTTTCTCCAAATTTCTGTAGACTAGCG

OA-11

TTCCCAAAGCCGAGTTTGCAGAAGTTTCCAAGTTAGTGACGGATTTGACCAAAG
TCCATACGGAATGTTGCCATGAGATTTGCTTGAATGTGC

OA-12

CTTCAGTTTACTGGAGATCGAATCTTGATTTTCACAGATATACTTGGCCAAGTCC
GCCCTGTCATCAGCACATTCAAGCAAATCTCC

OA-13

CGATCTCCAGTAACTGAAGGAATGTTGTTAAAAACCTCTTTTGGAAAAATCCC
ACTGCATTGCCGAAGTGGAACGATGAGATGCC

OA-14

CCTTTGCCTCAGCGTAGTTTTTGC AAACATCCTTACTTTCAACAAAATCAGCAGC
TAATGATGGCAAGTCAGCAGGCATCTCATCGTTTTCC

OA-15

CTACGCTGAGGCAAAGGATGTCTTCTTGGGCATGTTTTGTATGAATACGCAAG
AAGGCATCCAGATTACTCTGTCGTGCTGTTCCCCCGCTGA

OA-16

CCTATCGATACATTCATGAGGATCTGCAGCGGCACAGCACTTCTCTAGAGTGGTT
TCATATGTCTTGGCAAGTCTCAGCAACAGACCACA

然后按上文 1 (1) 中所述的方法经三次 PCR 扩增将此 8 个寡核苷酸 (OA-9 至 OA-16) 连接成带有 5'端 BamHI 和 3'端 ClaI 酶切位点, 长度为 564bp 的双链 HSA-II 片段. 该片段代表了 HSA 基因序列的中间部分. 同样, 按上文 1 (2) 和中所述方法纯化和克隆所合成的 HSA-II 片段, 并制备含有 HSA-II 的重组克隆载体 pBKS/HSA-II. 图 6 中图解显示了用于克隆和鉴定目的重组克隆载体 pBKS/HSA 的构建. 然后按上文 1 (3) 中所述的方法, 同样取 5 个携带 pBKS/HSA-II 的大肠杆菌克隆, 分离 DNA 并进行序列分析. 序列分析结果显示所分析的 5 个克隆中只有一个克隆的碱基序列带有单个碱基错误, 但该错误并未导致氨基酸序列的改变.

3.HSA-III 片段的合成与克隆

按照上文 1 (1) 所述的合成 HSA-I 片段的方法, 首先合成具有下示序列的 8 个寡核苷酸:

OA-17

CGCGGATCCGTGCCGCTGCAGATCCTCATGAATGTTATGCCAAAGTTTTTCGATGA
ATTTAAACCTCTTGTGGAAGAGCCACAAAATTTAATTAAACAAAATTGTGAG

OA-18

GGAGTTGACACTTGGGGTACTTTCTTGGTGTAAACGAACTAACAGCGCATTCTGG
AATTTGTATTACCAAGTTGCTCAAAAAGCTCACAATTTTGTTTAAT

OA-19

CCCCAAGTGTCAACTCCAACCTTTGGTAGAGGTCTCAAGAAACCTAGGTAAAGTG
GGTAGCAAGTGTTTGTAACATCCTGAAGCAAAAAGAATGC

OA-20

GACTCTGTCACCTTACTGGCGTTTTCTCATGCAACACACATAACTGGTTCAGGACA
ACGGATAGATAGTCTTCTGCACATGGCATTCTTTTTGCTTCAGG

OA-21

CCAGTAAGTGACAGAGTCACCAAGTGCTGCACAGAATCTTTGGTTAACAGGCCGA
CCATGCTTTTCAGCTCTGGAAGTCGATGAAACATACGTTCCCAAAGAGTTTAATG
C

OA-22

GTGTTTCACTAGCTCAACTAGTGCAGTTTGTTTCTTGATTTGTCTTTCCTTCTCAG
AAAGTGTACATATATCTGCATGGAAGGTGAAAGTTTCAGCATTAAACTCTTTGG
G

OA-23

GTTGAGCTAGTGAAACACAAGCCGAAGGCGACTAAAGAACAACCTGAAAGCTGT
TATGGATGGATGATTTTCGCGGCTTTTGTAGAAAAGTGTGTAATGCTGACG

OA-24

CCTATCGATCAAGAATTCCTATTATAAACCTAAGGCAGCTTGAGATGCAGCAAC
TAGTTTCTTACCTTCCTCGGCAAAGCAAGTTTCCTTATCGTCAGCCTTAGAAGAC

然后按上文 1 (1) 中所述方法经三次 PCR 扩增将此 8 个寡核苷酸链 (AO-17 至 OA-24) 连接成带有 5'端 BamHI 和 3'端 ClaI 酶切位点, 长度为 705bp 的双链 HSA-III 片段. 该片段代表了 HSA 基因序列的 3'端部分. 同

样按上述 1 (2) 中所述的相似方法纯化并克隆所合成的 HSA-III 片段, 并制备含有 HSA-III 的重组克隆载体 pBKS/HSA-III. 图 8 中图解显示了用于纯化、克隆和鉴定目的重组载体 pBKS/HSA-III 的构建. 然后按上文 1 (3) 中所述方法, 同样取 5 个携带 pBKS/HSA-III 的大肠杆菌克隆, 分离 DNA 并进行序列分析. 序列分析结果显示, 所有 5 个克隆均存在个别错误碱基. 为此, 按上文 1(3)中所述方法用带有正确碱基的亚片段取代带有错误碱基的片段. 在纠正 HSA-III 片段的碱基错误中, 借助核苷酸位置 216 处的 StyI 位点和核苷酸位置 373 处的 Hap-I 位点将三个亚片段重新连接成完整的 HSA-III 片段.

4. 全长度 HSA 基因的制备与克隆

基本上按照如上文所说的克隆 HSA 编码序列中各片段的方法制备携带全长 HSA 基因序列的重组克隆载体. 首先分别用 BamHI 和 HindIII, HindIII 和 PstI, 以及 PstI 和 ClaI 从相应的重组质粒 pBKS/HSA-I, pBKS/HSA-II 和 pBKS/HSA-III 中切出片段 HSA-I、HSA-II 和 HSA-III. 纯化各片段后, 将它们以相等的量混合在一起, 然后向所得各片段的混合物中加入用 BamHI 和 ClaI 双酶切成线性的质粒 pBKS. 图 12 显示了 HSA-I、HSA-II、HSA-III 片段和全长度 HSA DNA 片段的 1 % 琼脂糖凝胶电泳图谱. 按上述反应条件完成 4 个片段的连接, 重新环化并经筛选和序列分析后, 得到定名为 pBKS/HSA 的重组质粒. 进一步地全长度序列分析表明, 所得的 HSA 基因序列具有如图 1 所示的全长度核苷酸序列.

遗传稳定性实验结果显示, 按上述方法制得的重组质粒 pBKS/HSA 的 DH5 α 菌株在继式培养后, 所得 10 个单克隆中携带的质粒 DNA 具有与原始克隆完全相同的 DNA 序列, 表明本发明合成的 HSA 基因序列具有良好的遗传稳定性.

实施例 2: 重组表达载体的构建

本实施例描述用于在甲基营养型酵母中表达 HSA 的重组表达载体的构建. 一般说来, 可按本领域中已知的标准方法完成用于本发明目的的表达载体的构建 (如参见 Sambrook et al, Molecular Cloning, A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, 1989). 可按照欧洲专利申请 0344459 和国际专利申请 WO90/10697 中所述的方法构建用于表达本发明的全合成 HSA 基因的重组载体.

为了方便起见, 作为一种可供选择的方法, 我们利用可从市场上购得的适用于甲基营养型酵母系统的质粒 pPIC9 (Invitrogen) 构建用于分泌表达 HSA 的重组表达载体. 质粒 pPIC9 含有适于在甲基营养型酵母中表达和

分泌外源蛋白质的必要调控序列（非结构基因调控元件），其中包括 AOX1 启动子片段、 α 因子分泌信号（AMF 前原序列）、多克隆位点、3'AOX1 转录终止片段、HIS4 - ORF、3'AOX1 片段、ColE1 复制原点以及氨苄青霉素抗性基因。使用该载体还具有更便于检验本发明合成的 HSA 基因在甲基营养型酵母中分泌表达的优点。

为此，首先将携带质粒 pPIC19 的大肠杆菌 37 °C 振荡培养 15 小时，培养后离心（5000rpm，5 分钟）收集细胞沉淀物并以碱裂解法制备质粒 DNA（参见实施例 1 有关 pPKS 的制备方法），然后用 EcoRI 和 XhoI（各 3 单位）消化该质粒 DNA。在 1 % 琼脂糖凝胶上电泳分离并回收长度约 8.0Kb 的线性 DNA 片段。

将按上述方法（参见实施例 1）制备的 pBKS/HSA 质粒 DNA 与 2 μ l 多用酶切缓冲液（Promega）及 EcoRI 和 XhoI（各 3 单位）混合并于 37 °C 保温 1.5 小时后，在 1 % 琼脂糖凝胶上电泳分离并回收 HSA 基因片段。将大约 2 μ g（10 μ l）所得 HSA 基因片段与用同样酶切的 pPIC9 酶切产物混合，于 T4 DNA 连接酶（3 单位）存在下，在 16 °C 水浴中保温 17 小时，完成连接反应。

用所得到的连接混合物转化经过 0.1M CaCl₂ 处理的感受态大肠杆菌 DH5 α 细胞。然后在含有氨苄青霉素的 LB 平板上选择有氨苄青霉素抗性（Amp^R）的转化菌落，并从中提取质粒 DNA。用 BamHI 和 ClaI 消化所得质粒 DNA 并以 SDS - PAGE 法初步鉴定后，再经 DNA 序列分析筛选出在正确位置上以正确方向插入了含有 HSA 基因的重组质 pPIC9/HSA。

实施例 3：酵母宿主菌株的转化和阳性转化株的筛选

本实施例描述用上述方法制备的重组表达载体 pPIC9/HSA 转化巴斯德毕赤酵母细胞，以得到稳定表达和分泌 HSA 的重组体细胞的方法。

用碱裂解法从携带 pPIC9/HSA 质粒的大肠杆菌 DH5 α 中分离得到重组质粒 pPIC9/HSA。经苯酚/氯仿提取，并用无水乙醇沉淀后，得到经过纯化的可用于转化巴斯德毕赤酵母菌株的 pPIC9/HSA 的质粒 DNA。用 SalI（5 单位）于 37 °C 水浴中消化 10 μ g pPIC9/HSA，将其切成线性 DNA。直接用此酶切反应混合物转化巴斯德毕赤酵母 GS115（HIS⁻）菌株（NRRL Y-15851）。SalI 切割的 pPIC9/HSA 可整合到毕赤酵母染色体中的组氨酸脱氢酶基因中，而不破坏 AOX1 基因。由于 pPIC9/HSA 携带有 HIS 基因，故可基于 HIS⁺ 表型选择所得到的转化株。

将巴斯德毕赤酵母 GS115〔NRRL Y-15851〕接种于约 10ml YPD 培养基（1 % 酵母提取物，2 % 蛋白胨，2 % 葡萄糖）中，30 °C 振荡培养过

夜, 然后取 1ml 过夜培养物, 于 500ml 新鲜 YPD 培养基中培养约 12 - 14 小时, 至 $OD_{600} = 1.3-1.5$ 离心 (3000rpm, 5 分钟, 4 °C) 收集细胞, 并重新悬浮于用冰浴预冷的无菌水中. 用无菌水将细胞再洗涤两次后, 重新悬浮于 1M 山梨醇溶液中.

将 10 μ g 线性化的 pPIC9/HSA 质粒 DNA 加入 80 μ l 上述的 GS115 细胞悬液中, 混匀后, 转移到预冷却的电转移槽内, 冰浴 5 分钟. 使用电穿孔仪 (Invitrogen), 按仪器制造商推荐的条件 (电场强度 1500V/cm,) 10 毫秒以电穿孔法转化毕赤酵母细胞. 施加电脉冲后, 立即向电转移槽中加入 1ml 用冰浴预冷的 1M 山梨醇溶液, 并将内容物转移到微量离心管中. 取 200 μ l 细胞悬液铺敷于 MD 平板 (1.34 % YNB, 4×10^{-5} % 生物素, 0.5 % 甲醇) 和 MM (1.34 % YNB, 4×10^{-5} % 生物素, 1 % 右旋糖) 平板的相同位置上. 同时在 MD 和 MM 平板上分别接种携带 HSA 基因的 GS115-Albumin(HIS⁺ Mut^S) 半乳糖苷酶的 GS115(GS115- β -gal) 和携带 (HIS⁺ Mut⁺) 菌株作为对照以筛选具有 HIS⁺ Mut⁺ 表型的带有 pPIC9/HSA 重组质粒的阳性转化株. 为此将上述制备的平板于 30°C 培养两天, 其中具有 HIS⁺ Mut⁺ 表型的菌株在 MM 培养基和 MD 培养基上均能正常生长, 所得菌株大小与对照菌株相似. 具有 HIS⁺ Mut⁺ 表型的菌株能够在 MD 培养基上正常生长, 但在 MM 培养基上生长很慢或几乎不能生长.

将按上述方法筛选到的 50 个 HIS⁺ Mut⁺ 单克隆菌株分别接种 3ml YPD 液体培养基中, 30 °C 振荡培养 24 小时, 然后离心收集细胞. 将细胞沉淀重新悬浮于 120 μ l 无菌水中, 各加 0.4g 经无菌处理的玻璃珠, 剧烈振荡 3 分钟破碎细胞. 离心去除细胞沉淀后, 按前述方法依次用苯酚/氯仿 (1:1, 总体积 100 μ l) 和氯仿/异戊醇 (24:1, 总体积 100 μ l), 提取染色体 DNA.

可用 PCR 方法检测如上述制备的染色体 DNA, 以进一步筛选出携带有本发明合成的 HSA 基因的阳性转化株. 为此, 将 1 μ g (5 μ l) 上述被转化细胞的基因组 DNA 分别加入到由下列成分组成的 PCR 反应体系中:

10X PCR 反应缓冲液 (5 μ l), 基因组 DNA (1 μ g) 约 5 μ l; 100 mM dNTPs (1 μ l); 5' AOX1 引物 (0.1 μ g/ μ l 约 5 μ l); 3' AOX1 引物 (0.1 μ g/ μ l 约 5 μ l); 补加无菌水至总体积为 50 μ l. 再加入 0.25 μ l Taq DNA 聚合酶 (5U/ μ l), 其中 5' AOX1 引物, 3' AOX1 引物均购自 Invitrogen 公司.

PCR 反应条件是: 94 °C 热变性 2 分钟, 然后按下述条件进行 25 次循环: 94 °C 变性 (40 秒), 55 °C 复性 (90 秒), 72 °C 延伸 (1 分钟), 最后 72 °C 延伸 7 分钟. PCR 反应完成后, 将 PCR 反应产物于 1 % 琼脂糖凝胶电泳上电泳, 观察是否有扩增的 HSA 基因片段, 以进一步筛选出含有 HSA 基因的转化株. 该细胞株已按照用于专利程序的国际公认微生物保藏

布达佩斯条约的规定, 携带重组质粒 pPIC9/HSA 的毕赤氏酵母 GS115 菌株将保藏于中国微生物菌种保藏管理委员会普通微生物保藏中心(CGMCC), 其保藏登记号为 CGMCC NO. 0354。

实施例4: HSA 基因在 GS115 酵母中的表达及蛋白质产物的免疫印迹分析

(1500rpm)使罐内溶解氧保持在20%以上空气饱和度。在整个发酵过程中,定时加入市售的Struktol J673消泡剂以控制过多的气泡生成。当甘油逐渐消耗后,向罐内泵入含0.5%甲醇但不含甘油的BMMY培养基(700ml),30℃下继续进行补料分批培养。此培养阶段内发酵罐中的甲醇浓度始终保持在0.3% - 0.7%。发酵后20小时开始收集培养物,以天然HSA标准品作为对照,经10%聚丙烯酰胺凝胶电泳并染色后检测光密度,估测发酵培养物HSA的浓度为4.5 g/L。发酵完成后,离心(10,000rpm,15分钟)收集培养物上清液,于70℃加热处理20分钟,然后快速冷却到15℃,离心(15,000rpm,20分钟)去除包括微量蛋白酶在内的热变性杂质后使离心上清液通过孔径为0.45 μm的微孔滤膜(Millipore)过滤。过滤后再用截留分子量为30,000的超滤膜(Millipore)将此上清液浓缩至大约0.5L体积。然后用乙酸将溶液pH调节到4.5左右,通过预先用50mM醋酸钠缓冲液(pH4.5)平衡过的RBO2脱色树脂(5 × 25cm,柱床体积约500ml)进行脱色,向所得澄清透明洗出液内加入硫酸铵,使其终浓度为10%。然后经预先用50mM磷酸盐 + 10%(NH₄)₂SO₄, pH6.8缓冲液平衡过的phenyl Sepharose柱吸附,用一倍柱床体积平衡的上述预平衡缓冲液洗涤柱,最后用洗脱缓冲液(200mM PB + 30%异丙醇缓冲液, pH6.8)洗脱并收集280nm吸收峰的洗脱部分。进一步将该洗脱液经G-25凝胶层析柱脱盐后得到基本上适于临床应用的高纯度HSA产物。

序列表

(I)、一般信息

(I) 申请人: 上海海济生物工程有限公司

(II) 发明名称: 重组人血清白蛋白的生产方法

(III) 序列数: 2

(IV) 通讯地址:

(A) 联系人: 李绍极
 (B) 街道: 太原路 192 号
 (C) 城市: 上海
 (D) 国家: 中华人民共和国
 (E) 邮编: 200031

(V) 计算机可读形式:

(A) 介质类型: 3.5 英寸软盘
 (B) 计算机: AST Pentium III+4/66d
 (C) 操作系统: Windows 95
 (D) 软件: Word 97

(VI) 电讯信息:

(A) 电话: 86 - 21 - 64735898
 (B) 电传: 86 - 21 - 64674742

(2) SEQ ID NO: 1 的信息

(I) 序列特征:

(A) 长度: 1758 个碱基对
 (B) 类型: 核酸
 (C) 链型: 双链
 (D) 拓扑结构: 线性

(II) 分子类型: DNA

(III) 序列描述: SEQ ID NO: 1

1
 GATGCACACA AGAGTGAGGT GCTCATCGG TTAAAGATT TGGGTGAGGA
 51
 AAATTTCAAA GCCTTAGTGT TAATTGCTTT TGCCCAATAT CTTAGCAAT
 101
 GTCCATTGA GGATCATGTA AAGTTGGTTA ACGAGGTCAC TGAGTTCGCT
 151
 AAAACTTGTG TTGCAGACGA ATCTGCTGAA AATTGTGATA AGTCTCTGCA
 201
 TACCCTATTT GGTGACAAAT TGTGCACAGT CGCTACCCTT CGAGAACTT
 251
 ACGGTGAAAT GGCCGACTGC TGTGCAAAGC AAGAACCTGA GAGAAACGAA
 301
 TGCTTTTTGC AGCACAAAGA TGACAACCCT AATCTTCCAC GTTAGTGAG
 351
 ACCTGAAGTT GATGTAATGT GTACAGCCTT CCATGATAAT GAAGAACTT
 401
 TTTTGAAGAA GTACTTATAT GAAATTGCAA GAAGGCACCC ATACTTCTAT
 451
 GCCCCGGAGC TACTGTTCTT TGCTAAACGT TATAAGCTG CATTACAGA
 501
 ATGTTGCCAA GCTGCCGATA AGGCAGCTTG TCTGTTGCCA AAGCTAGACG
 551
 AATTAAGAGA TGAGGGTAAA GCTTCTTCTG CCAAACAAAG ATTGAAGTGC
 601
 GCTAGTCTAC AGAAATTTGG AGAAAGAGCT TTAAGGCAT GGGCTGTAGC
 651
 GCGCCTGAGC CAGAGATTTT CCAAAGCCGA GTTTGCAGAA GTTTCCAAGT
 701
 TAGTGACGGA TTTGACCAAA GTCCATACGG AATGTTGCCA TGGAGATTTG
 751
 CTTGAATGTG CTGATGACAG GGCGGACTTG GCCAAGTATA TCTGTGAAAA
 801
 TCAAGATTCTG ATCTCCAGTA AACTGAAGGA ATGTTGTGAA AAACCTCTTT
 851
 TGGAAAAATC CCACTGCATT GCCGAAGTGG AAAACGATGA GATGCCTGCT
 901
 GACTTGCCAT CATTAGCTGC TGATTTTGTT GAAAGTAAGG ATGTTTGCAA
 951
 AAACACGCT GAGGCAAAGG ATGTCTTCTT GGGCATGTTT TTGTATGAAT
 1001

ACGCAAGAAG GCATCCAGAT TACTCTGTCG TGCTGTTGCT GAGACTTGCC
 1051
 AAGACATATG AAACCACTCT AGAGAAGTGC TGTGCCGCTG CAGATCCTCA

1101
 TGAATGTTAT GCCAAAGTTT TCGATGAATT TAAACCTCTT GTGGAAGAGC
 1151
 CACAAAATTT AATTAAACAA AATTGTGAGC TTTTGTAGCA ACTTGGTGAA
 1201
 TACAAATTCC AGAATGCGCT GTTAGTTCGT TACACCAAGA AAGTACCCCA
 1251
 AGTGTCAACT CCAACTTTGG TAGAGGTCTC AAGAAACCTA GGTAAGTGG
 1301
 GTACCAAGTG TTGTAAACAT CCTGAAGCAA AAAGAATGCC ATGTGCAGAA
 1351
 GACTATCTAT CCGTTGTCCT GAACCAGTTA TGTGTGTTGC ATGAGAAAAC
 1401
 GCCAGTAAGT GACAGAGTCA CCAAGTGCTG CACAGAATCT TTGGTTAACA
 1451
 GGCGACCATG CTTTTCAGCT CTGGAAGTCG ATGAAACATA CGTCCCAAA
 1501
 GAGTTTAATG CTGAACTTT CACCTTCCAT GCAGATATAT GTACACTTTC
 1551
 TGAGAAGGAA AGACAAATCA AGAAACAAAC TGCACTAGTT GAGCTAGTGA
 1601
 AACACAAGCC GAAGGCGACT AAAGAACAAC TGAAAGCTGT TATGGATGAT
 1651
 TTCGCGGCTT TTGTAGAAAA GTGTTGTAAG GCTGACGATA AGGAAACTTG
 1701
 CTTTGCCGAG GAAGGTAAGA AACTAGTTGC TGCATCTCAA GCTGCCTTAG
 1751
 GTTTATAA

(3) SEQ ID NO: 2 的信息

(I) 序列特征:

(A) 长度: 585 个氨基酸
 (B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(II) 分子类型: 蛋白质

(III) 序列描述: SEQ ID NO: 2

1
 Asp Ala His Lys Ser Glu Val Ala His Arg Phe Lys Asp Leu Gly Glu Glu
 18
 Asn Phe Lys Ala Leu Val Leu Ile Ala Phe Ala Gln Tyr Leu Gln Gln Cys
 35
 pro Phe Glu Asp His Val Lys Leu Val Asn Glu Val Thr Glu Phe Ala Lys
 52
 Thr Cys Val Ala Asp Glu Ser Ala Glu Asn Cys Asp Lys Ser Leu His Thr
 69
 Leu Phe Gly Asp Lys Leu Cys Thr Val Ala Thr Leu Arg Glu Thr Tyr Gly
 86
 Glu Met Ala Asp Cys Cys Ala Lys Gln Glu Pro Glu Arg Asn Glu Cys Phe
 103
 Leu Gln His Lys Asp Asp Asn Pro Asn Leu Pro Arg Leu Val Arg Pro Glu
 120
 Val Asp Val Met Cys Thr Ala Phe His Asp Asn Glu Glu Thr Phe Leu Lys
 137
 Lys Tyr Leu Tyr Glu Ile Ala Arg Arg His Pro Tyr Phe Tyr Ala Pro Glu
 154
 Leu Leu Phe Phe Ala Lys Arg Tyr Lys Ala Ala Phe Thr Glu Cys Cys Gln
 171
 Ala Ala Asp Lys Ala Ala Cys Leu Leu Pro Lys Leu Asp Glu Leu Arg Asp
 188
 Glu Gly Lys Ala Ser Ser Ala Lys Gln Arg Leu Lys Cys Ala Ser Leu Gln
 205
 Lys Phe Gly Glu Arg Ala Phe Lys Ala Trp Ala Val Ala Arg Leu Ser Gln
 222
 Arg Phe Pro Lys Ala Glu Phe Ala Glu Val Ser Lys Leu Val Thr Asp Leu
 239
 Thr Lys Val His Thr Glu Cys Cys His Gly Asp Leu Leu Glu Cys Ala Asp
 256
 Asp Arg Ala Asp Leu Ala Lys Tyr Ile Cys Glu Asn Gln Asp Ser Ile Ser

273

Ser Lys Leu Lys Glu Cys Cys Glu Lys Pro Leu Leu Glu Lys Ser His Cys

290

Ile Ala Glu Val Glu Asn Asp Glu Met Pro Ala Asp Leu Pro Ser Leu Ala

307

Ala Asp Phe Val Glu Ser Lys Asp Val Cys Lys Asn Tyr Ala Glu Ala Lys

324

Asp Val Phe Leu Gly Met Phe Leu Tyr Glu Tyr Ala Arg Arg His Pro Asp

341

Tyr Ser Val Val Leu Leu Leu Arg Leu Ala Lys Thr Tyr Glu Thr Thr Leu

358

Glu Lys Cys Cys Ala Ala Ala Asp Pro His Glu Cys Tyr Ala Lys Val Phe

375

Asp Glu Phe Lys Pro Leu Val Glu Glu Pro Gln Asn Leu Ile Lys Gln Asn

392

Cys Glu Leu Phe Glu Gln Leu Gly Glu Tyr Lys Phe Gln Asn Ala Leu Leu

409

Val Arg Tyr Thr Lys Lys Val Pro Gln Val Ser Thr Pro Thr Leu Val Glu

426

Val Ser Arg Asn Leu Gly Lys Val Gly Ser Lys Cys Cys Lys His Pro Glu

443

Ala Lys Arg Met Pro Cys Ala Glu Asp Tyr Leu Ser Val Val Leu Asn Gln

460

Leu Cys Val Leu His Glu Lys Thr Pro Val Ser Asp Arg Val Thr Lys Cys

477

Cys Thr Glu Ser Leu Val Asn Arg Arg Pro Cys Phe Ser Ala Leu Glu Val

494

Asp Glu Thr Tyr Val Pro Lys Glu Phe Asn Ala Glu Thr Phe Thr Phe His

511

Ala Asp Ile Cys Thr Leu Ser Glu Lys Glu Arg Gln Ile Lys Lys Gln Thr

528

Ala Leu Val Glu Leu Val Lys His Lys Pro Lys Ala Thr Lys Glu Gln Leu

545

Lys Ala Val Met Asp Asp Phe Ala Ala Phe Val Glu Lys cys Cys Lys Ala

562

Asp Asp Lys Glu Thr Cys Phe Ala Glu Glu Gly Lys Lys Leu Val Ala Ala

579

585

Ser Gln Ala Ala Leu Gly Leu

说 明 书 附 图

A.

[illegible]

B.

10 Jan I	201 Oct II	424 Dec I	600 Nov I	1110 Dec II	1440 Dec III	1789 Oct I
10 Jan II	243 Sep I	465 Dec I	670 Dec III	1100 Oct I	1470 Oct I	1783 Dec I
10 Jan I	243 Sep I 2001	505 Oct I	789 Dec I	1100 Nov II	1432 Oct I	1788 Dec I
4 Oct II	243 Apr I	431 Sep I	828 Oct I	1009 Nov I	1070 Dec I	
			800 Oct I	1001 Nov I	1004 Nov 7624 I	
					1308 Apr II	

2

1/1	31/11
GAG GAA TTC CTT GAG AAA AGA GAT GAA GAC	AAG ACT GAG GTT GCT CAT GAG GAT TTT AAA GAT
gly gly ser leu glu lya arg asp ala his	lys ser glu val ala his arg phe lya asp
61/21	91/31
TTG GGT GAG GAA AAT TTC AAA GCC TTA GTG	TTA ATT GCT TTT GCC CAA TAT CTT CAG CAA
leu gly glu glu asn phe lys ala leu val	leu ile ala phe ala glu tyr leu gln gln
121/41	151/51
TGT CCA TTT GAG GAT CAT GTA AAG TTG GTT	AAC GAG GTC ACT GAG TTC GCT AAA ACT TGT
cys pro phe glu asp his val lys leu val	asn glu val thr glu phe ala lys thr cys
181/61	211/71
GTT GCA GAC GAA TCT GCT GAA AAT TGT GAT	AAG TCT CTT CAT ACC CTA TTT GGT GAC AAA
val ala asp glu ser ala glu asn cys asp	lys ser leu his thr leu phe gly asp lys
241/81	271/91
TTG TGC ACA GTC GCT ACC CTT CGA GAA ACT	TAC GGT GAA ATG GCC GAC TGC TGT GCA AAG
leu cys thr val ala thr leu arg glu thr	tyr gly glu met ala asp cys cys ala lys
301/101	331/111
CAA GAA CCT GAG ACA AAC GAA TTC TTT TTG	CAG CAC AAA GAT GAC AAC CCT AAT CTT CCA
gln glu pro glu arg asn glu cys phe leu	gln his lys asp asp asn pro asn leu pro
361/121	391/131
GCT TTA GTG AGA CCT GAA GTT GAT GTG ATG	TGT ACA GCC TTC CAT GAT AAT GAA GAA ACT
arg leu val arg pro glu val asp val met	cys thr ala phe his asp asn glu glu thr
421/141	451/151
TTT TTG AAG AAG TAC TTA TAT GAA AAT GCA	AGA AGG CAC CCA TAC TTC TAT GCC CCG GAG
phe leu lys lys tyr leu tyr glu ile ala	arg arg his pro tyr phe tyr ala pro glu
481/161	511/171
CCT CTT TTC TTT GCT AAA GGT TAT AAA GCT	GCA TTC ACA GAA TGT TTC CAA GCT GCC GAT
leu leu phe phe ala lys arg tyr lys ala	ala phe thr glu cys cys gln ala ala asp
541/181	571/191
AAG GCA GCT TGT CTT TTG CCA AAG CTA GAC	GAA TTA AGA GAT GAG GGT AAA GCT TCT TCT
lys ala ala cys leu leu pro lys leu asp	glu leu arg asp glu gly lys ala ser ser
601/201	
GCC ATC GAT AGG	
ala ile asp arg	

图 3

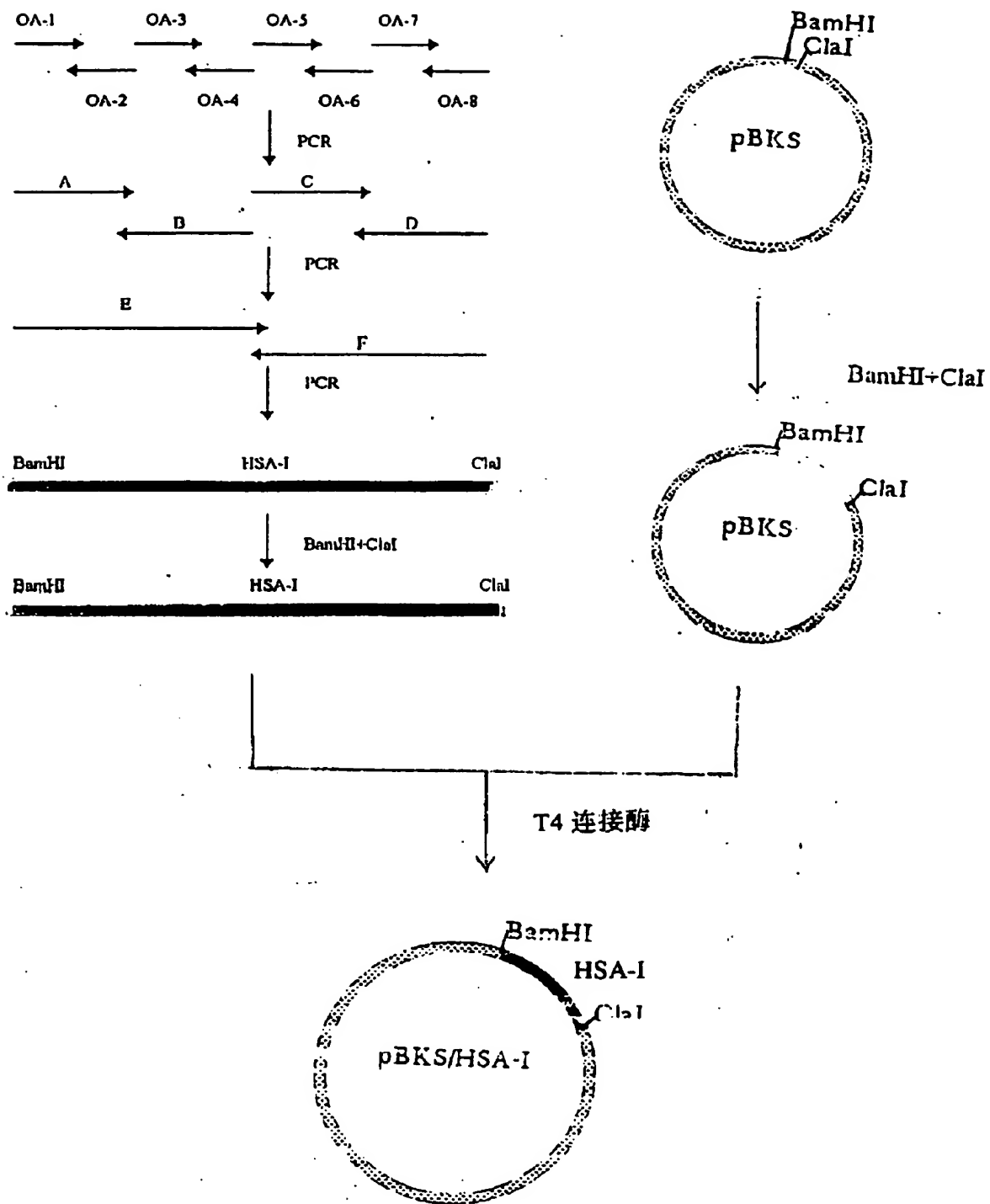


图 4

1/1	GGC GGA TCC GAG GGT AAA GCT TCT TCT GCC	31/11	AAA CAA ACA TTG AAG TGC GCT AGT CTA CAG
gly gly ser glu gly lys ala ser ser ala	lys gln arg leu lys cys ala ser leu gln		
61/21	AAA TTT GCA GAA ACA GCT TTT AAG GCA TTT	91/31	GCT CTA GCG CGC CTG ACC CAG ACA TTT CCC
lys phe gly glu arg ala phe lys ala tyr	ala val ala arg leu ser glu arg phe pro		
121/41	AAA GCC GAG TTT GCA GAA GTT TCC AAG TTA	151/51	CTG ACC GAT TTG ACC AAA GTC CAT ACC GAA
lys ala glu phe ala glu val ser lys leu	val thr asp leu thr lys val his thr glu		
181/61	TTT TGC CAT GGA GAT TTT CTT GAA TTT GCT	211/71	GAT GAC AAG GCG GAC TTG GCC AAG TAT ATC
cys cys his gly asp leu leu glu cys ala	asp asp arg ala asp leu ala lys tyr ile		
241/81	TTT GAA AAT CAA GAT TCG ATC TCC AGT AAA	271/91	CTG AAG GAA TGT TGT GAA AAA CCT CTT TTG
cys glu asn gln asp ser ile ser ser lys	leu lys glu cys cys glu lys pro leu leu		
301/101	GAA AAA TCC CAC TGC ATT GCC GAA GTG GAA	331/111	AAC GAT GAG ATG CCT GCT GAC TTG CCA TCA
glu lys ser his cys ile ala glu val glu	asn asp glu met pro ala asp leu pro ser		
361/121	TTA GCT GCT GAT TTT GTT GAA AGT AAG GAT	391/131	TTT TGC AAA AAC TAC GCT GAG GCA AAG GAT
leu ala ala asp phe val glu ser lys asp	val cys lys asn tyr ala glu ala lys asp		
421/141	GTC TTC TTG GGC ATG TTT TTG TAT GAA TAC	451/151	GCA AGA ACG CAT CCA GAT TAC TCT GTC GTG
val phe leu gly met phe leu tyr glu tyr	ala arg arg his pro asp tyr ser val val		
481/161	CTG TTG CTG AGA CTT GCC AAG ACA TAT GAA	511/171	ACC ACT CTA GAG AAG TGC TGT GCC GCT GCA
leu leu leu arg leu ala lys thr tyr glu	thr thr leu glu lys cys cys ala ala ala		
541/181	GAT GCT CAT GAA TTT ATC GAT AAG		
asp pro his glu cys ile asp arg			

图 5

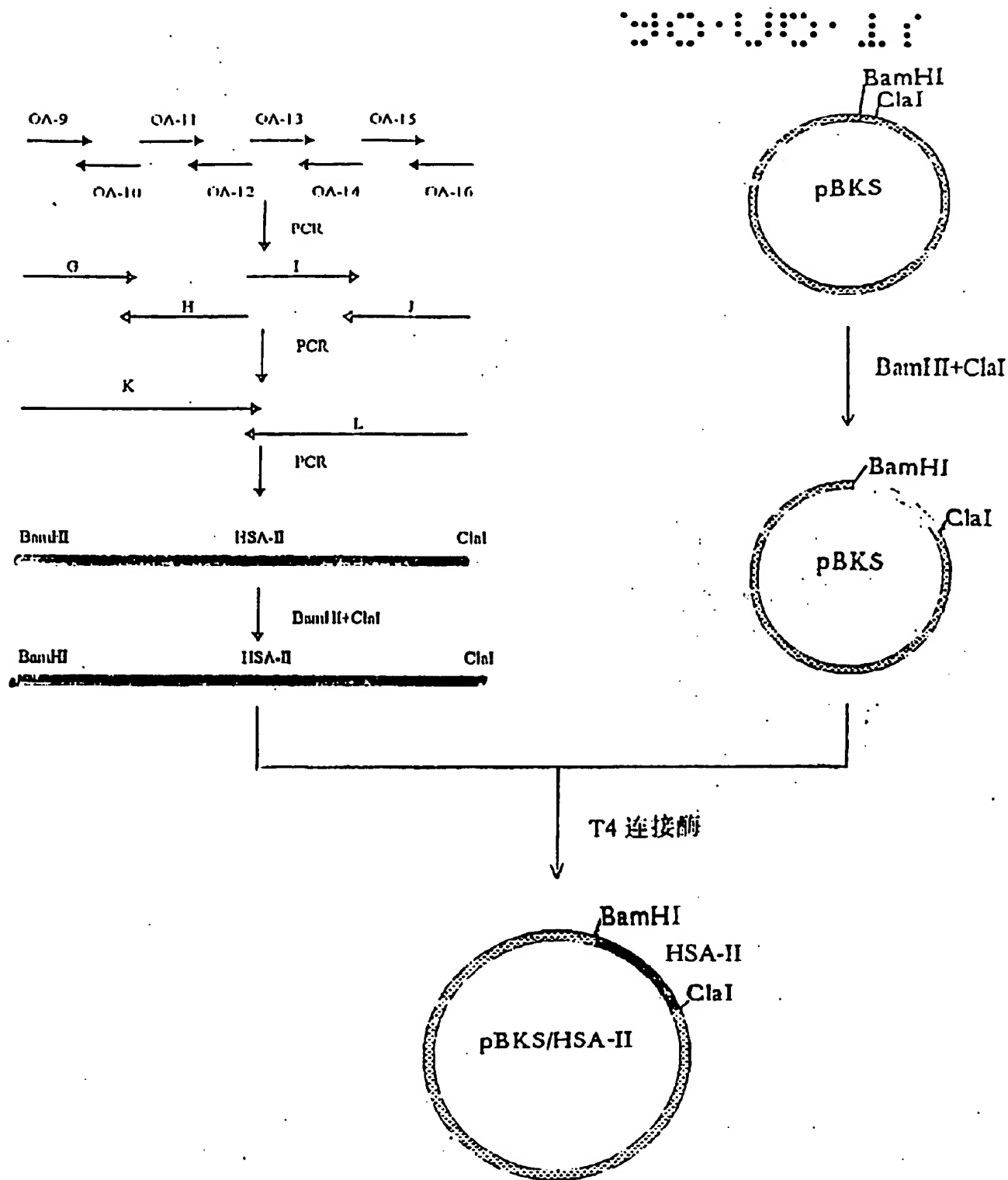


图 6

1/1 31/11
 GAG GAA TTT TGT GAT GAT GAA GAT GAT GAT GAA TTT TAT GTC AAA GGT TTC GAT GAA TTT
 gly gly ser cys ala ala ala asp pro his glu cys tyr ala lys val phe asp glu phe
 61/21 91/31
 AAA CCT CTT GTG GAA GAG CCA CAA AAT TTA ATT AAA CAA AAT TGT GAG CTT TTT GAG CAA
 lys pro leu val glu glu pro gln asn leu ile lys gln asn cys glu leu phe glu gln
 121/41 151/51
 CTT GGT GAA TAC AAA TTC GAG AAT GCG CTG TTA GTT GGT TAC ACC AAG AAA GTA CCC CAA
 leu gly glu tyr lys phe gln asn ala leu leu val arg tyr thr lys lys val pro gln
 181/61 211/71
 GTG TCA ACT CCA ACT TTT GTA GAG GTC TCA AGA AAC GTA GGT AAA GTG GGT ACC AAG TGT
 val ser thr pro thr leu val glu val ser arg asn leu gly lys val gly ser lys cys
 241/81 271/91
 TGT AAA CAT CCT GAA GCA AAA AGA ATG CCA TGT GCA GAA GAC TAT CTA TCC GTT GTC CTG
 cys lys his pro glu ala lys arg met pro cys ala glu asp tyr leu ser val val leu
 301/101 331/111
 AAC CAG TTA TGT GTG TTG CAT GAG AAA ACC CCA GTA AGT GAC AGA GTC ACC AAG TGC TGC
 asn gln leu cys val leu his glu lys thr pro val ser asp arg val thr lys cys cys
 361/121 391/131
 ACA GAA TCT TTG GTT AAC AGG CGA CCA TGC TTT TCA GCT CTG GAA GTC GAT GAA ACA TAC
 thr glu ser leu val asn arg arg pro cys phe ser ala leu glu val asp glu thr tyr
 421/141 451/151
 GTT CCC AAA GAG TTT AAT GAT GAA ACT TTC ATT TTC CAT GCA GAT ATA TGT ACA CTT TCT
 val pro lys glu phe asn ala glu thr phe thr phe his ala asp ile cys thr leu ser
 481/161 511/171
 GAG AAG GAA AGA CAA ATC AAG AAA CAA ACT GCA CTA GTT GAG CTA GTG AAA CAC AAG CCG
 glu lys glu arg gln ile lys lys gln thr ala leu val glu leu val lys his lys pro
 541/181 571/191
 AAG GCG ACT AAA GAA CAA CTG AAA GCT GTT AIG GAT GAT TTC GCG GCT TTT GTA GAA AAG
 lys ala thr lys glu gln leu lys ala val met asp asp phe ala ala phe val glu lys
 601/201 631/211
 TTT TTT AAG GGT GAC GAT AAG GAA ACT TTC TTT GCG GAG GAA GAT AAG AAA CTA GTT GCT
 cys cys lys ala asp asp lys glu thr cys phe ala glu glu gly lys lys leu val ala
 661/221 691/231
 CCA TCT CAA GCT GCC TTA GGT TTA TAA TAG GAA TTC ATC GAT AAG
 ala ser gln ala ala leu gly leu och amb glu phe ile asp arg

图 7

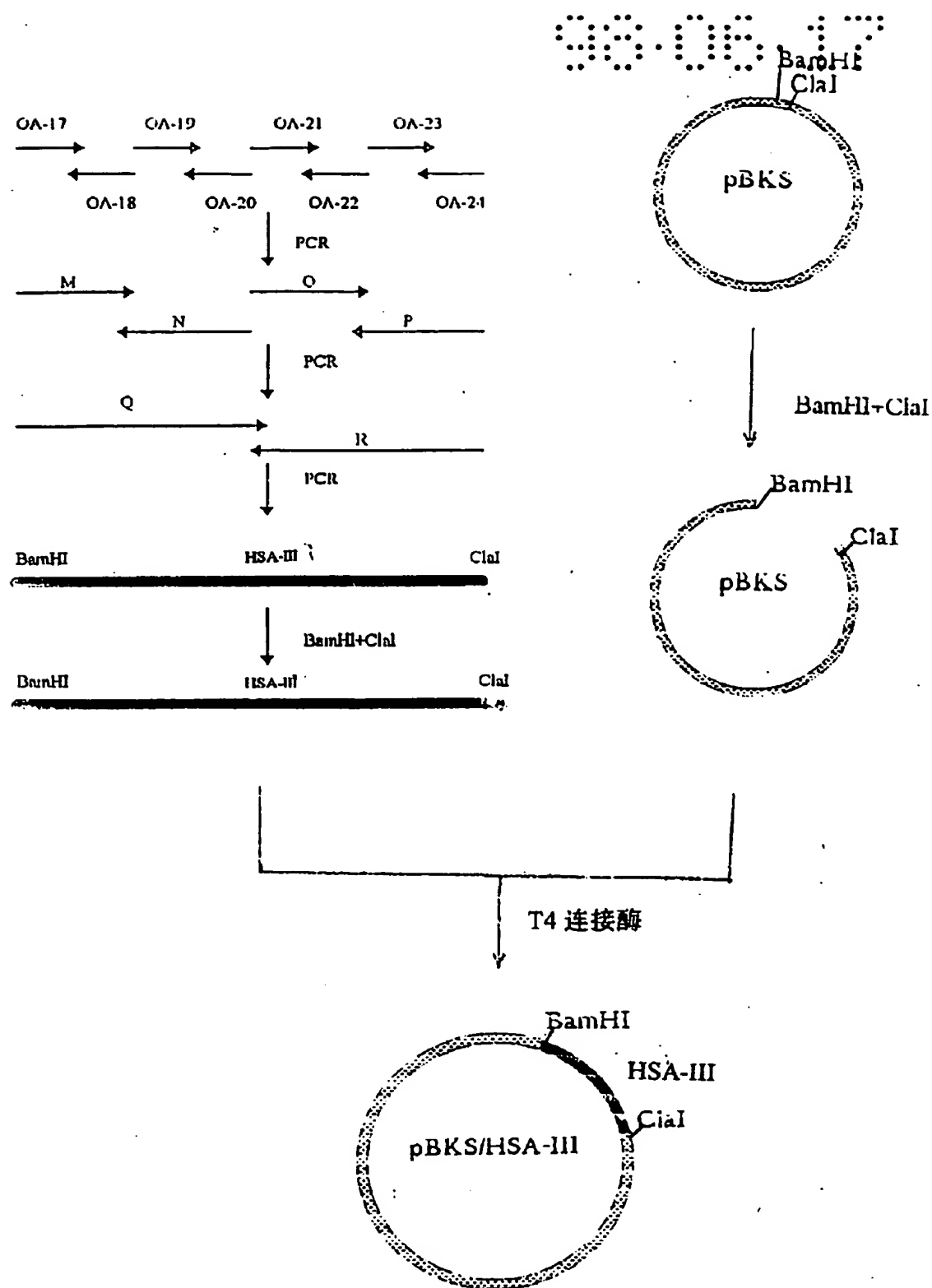


图 8

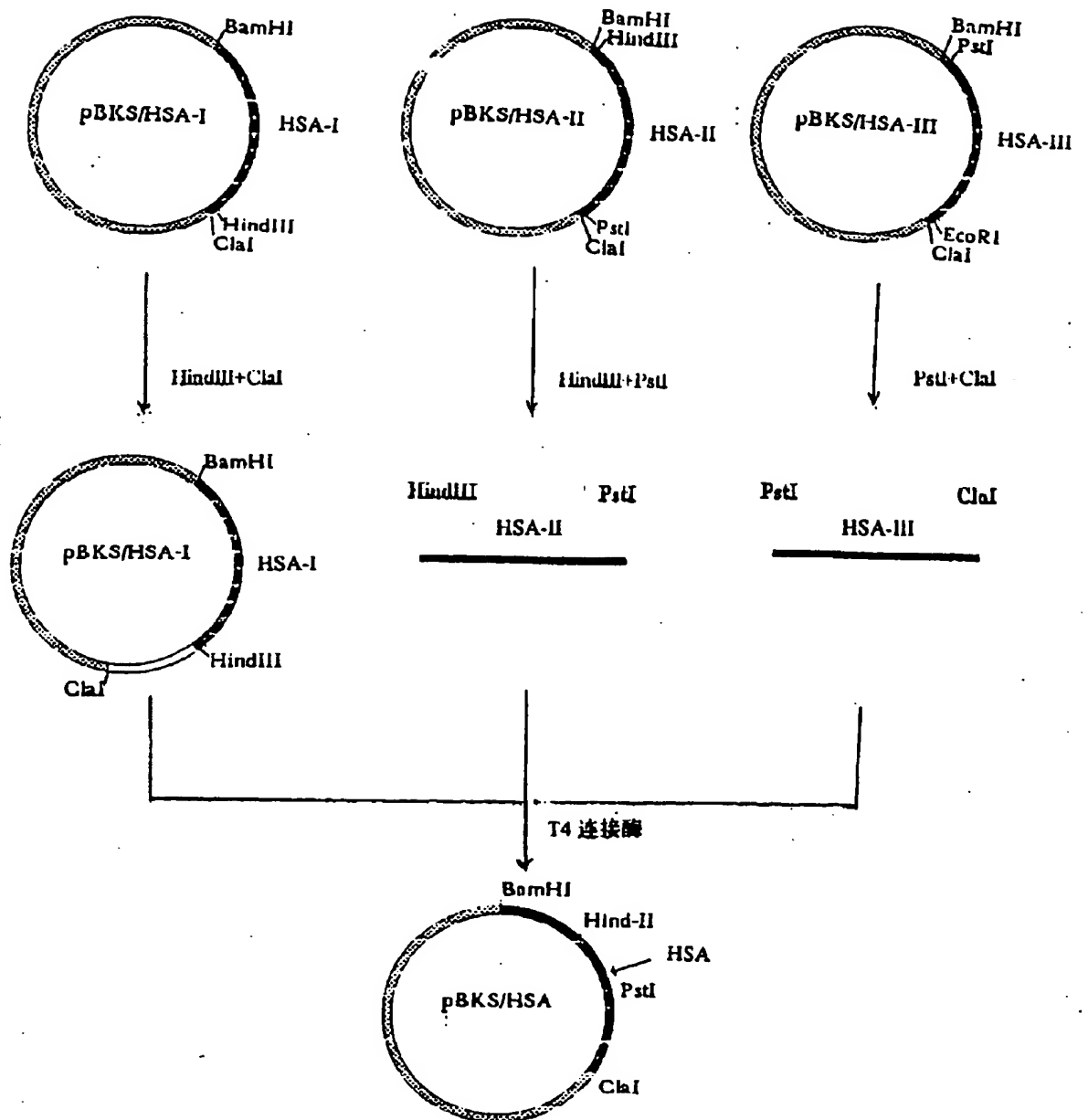


图 9

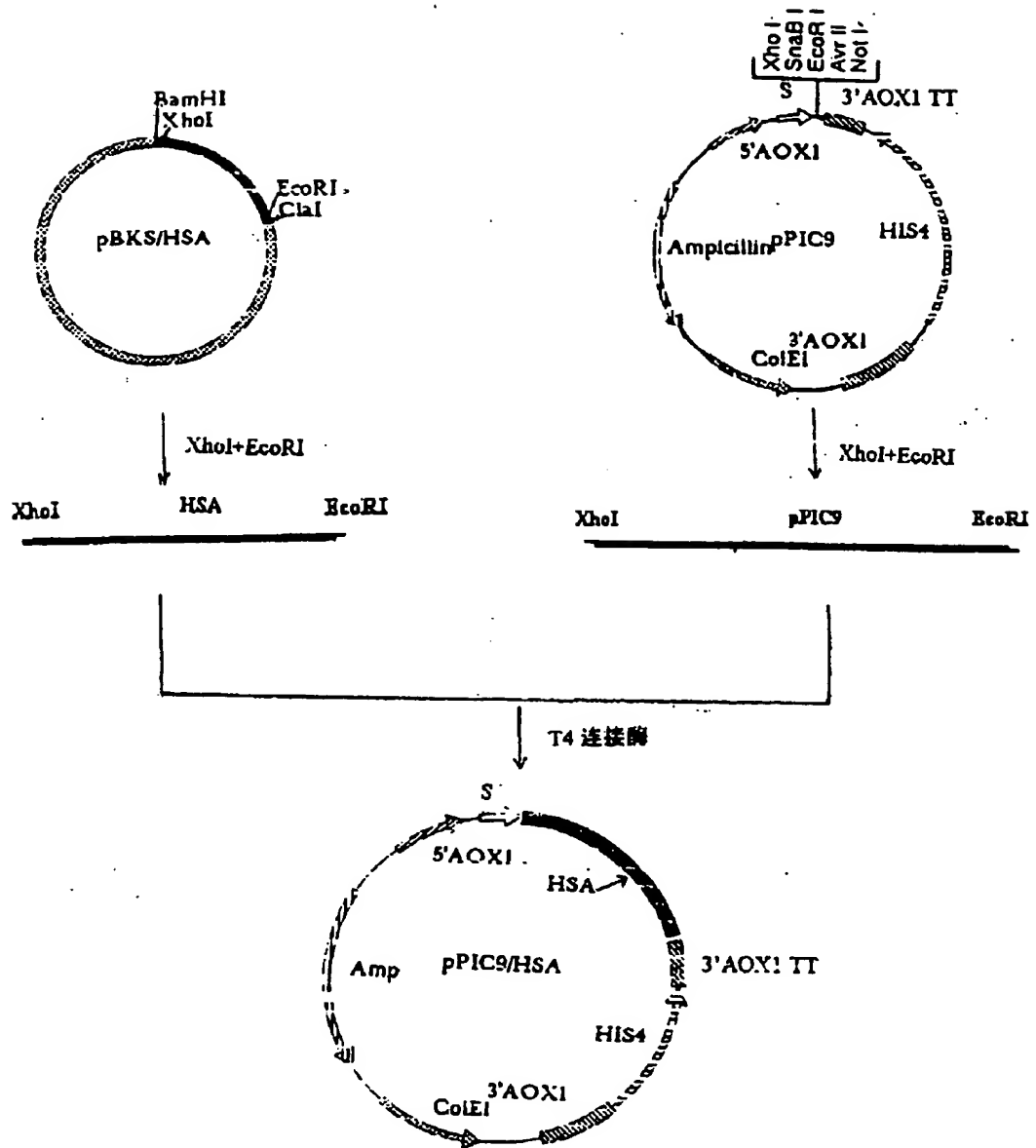


图 10

Aox1 mRNA 5'末端

TATCATCATT ATTAGCTTAC TTTCATAATT GCGACTGGTT CCAATTGACA
AGCTTTTGAT TTTAACGACTTTTAACGACAACCTTGAGAAGATCAAAAAACAATAATTAT

α-因子序列

TCGAAGGATCCAAACG ATG AGA TTT CCT TCA ATT TTT ACT GCA GTT TTA TTC
GCA GCA TCC TCC GCA TTA GCT GCT CCA GTC AAC ACT ACA ACA GAA GAT GAA
ACG GCA CAA ATT CCG GCT GAA GCT GTC ATC GGT TAC TCA GAT TTA GAA GGG
GAT TTC GAT GTT GCT GTT TTG CCA TTT TCC AAC AGC ACA AAT AAC GGG TTA TTG
TTT ATA AAT ACT ACT ATT GCC AGC ATT GCT GCT AAA GAA GAA GGG GTA TCT

XhoI

信号肽切位点

CTC GAG AAA AGA GAT GCA

Leu Glu Lys Arg Asp Ala

α-因子信号肽

+ 1

HSA

←—————→

EcoRI

AOX1 mRNA 3'末端

TTA TAA TAG AA TTC CCT AGG GCG GCC GCG AAT TAA TTTTTTGTCA.

Leu OCH AMB

图 11

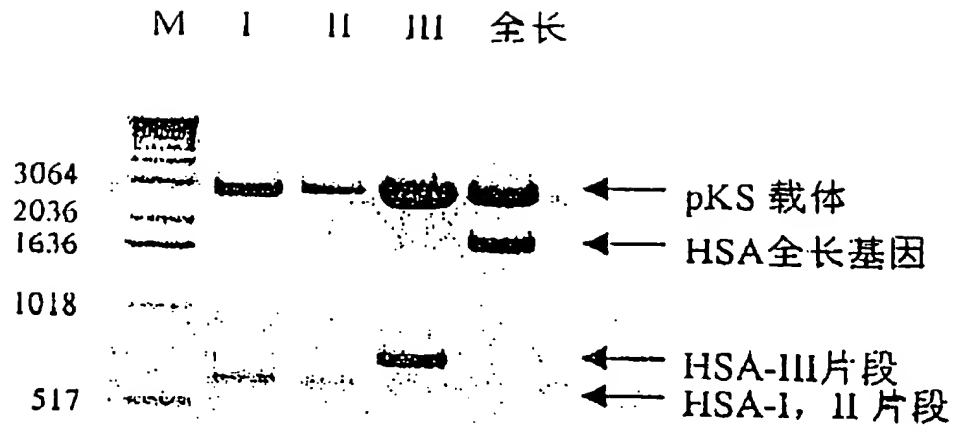


图 12

98.06.17

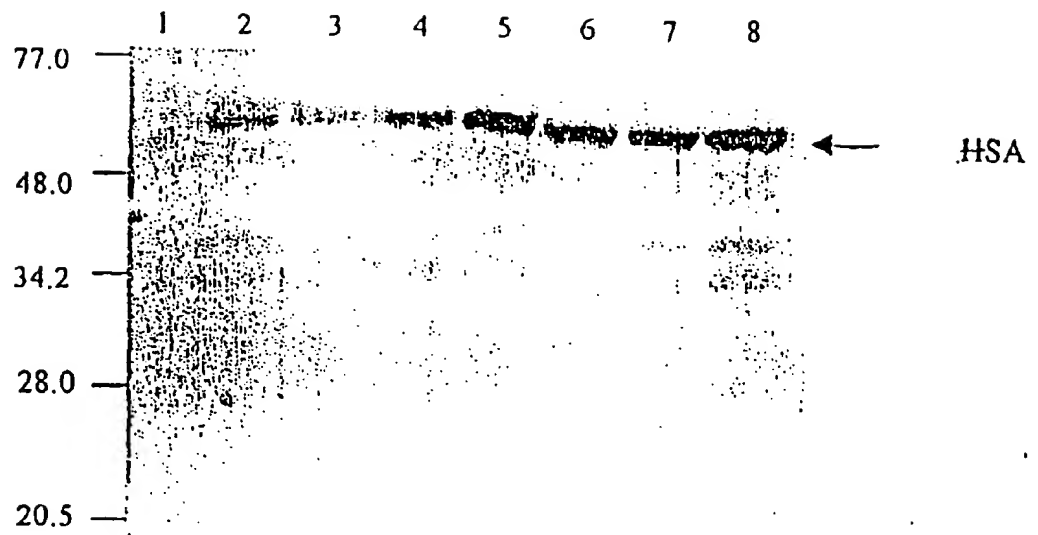


图 13

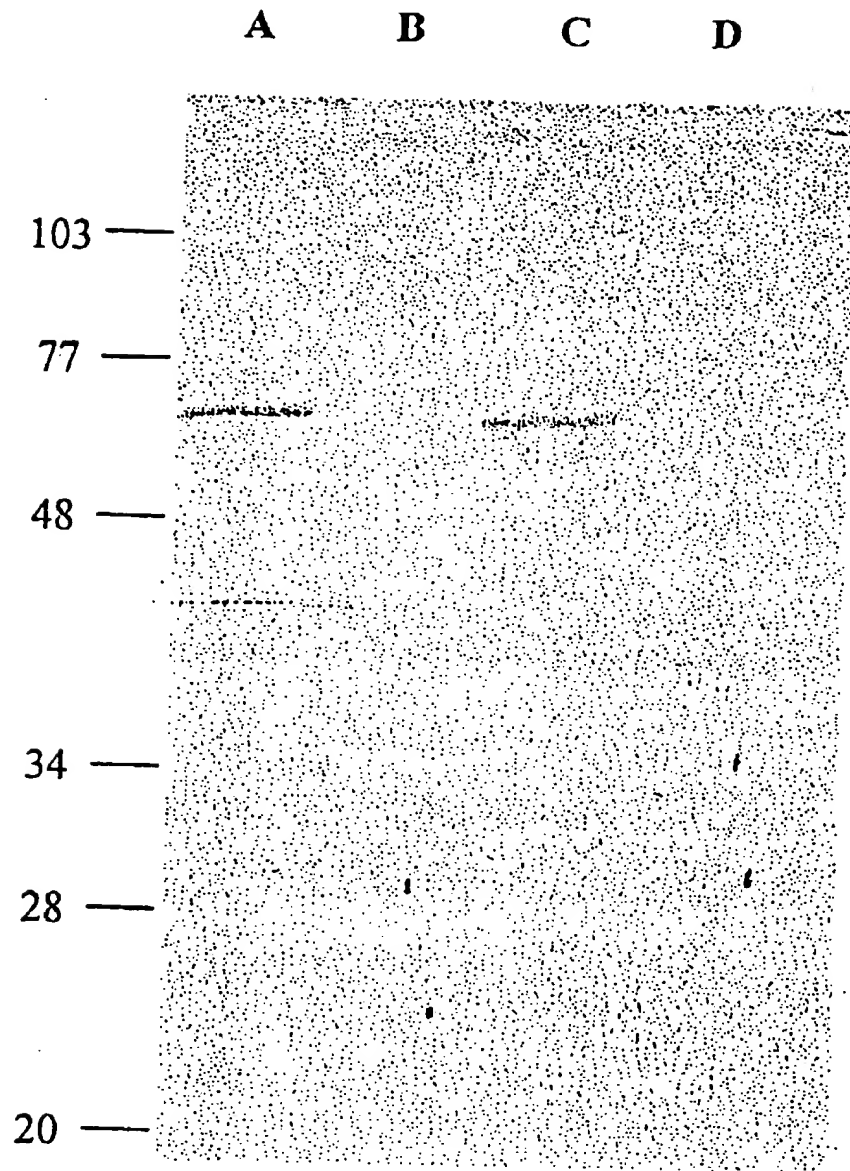


图 14